Lecture notes on

sub-Riemannian geometry

by Enrico Le Donne.

http://www.math.ethz.ch/~ledonnee/

Contents

Chapter 0

A brief introduction

These notes are mostly based on the following books, papers, and preprints: [CG90, Bel96, Gro99, AFP00, BBI01, Mon02, Bul02, AKLD08]. They have been written for the course 'Sub-Riemannian Geometry' at ETH in Zurich during Fall 2009.

Sub-Riemannian geometry is a generalization of Riemannian geometry. Roughly speaking, a sub-Riemannian manifold is a Riemannian manifold together with a constrain on admissible direction of movements. In Riemannian geometry any (smoothly embedded) curve has locally finite length. In sub-Riemannian geometry, if the curve fails to satisfy the obligation of the constrain, then it would have infinite length.

One classical example one should carry in mind is coming from mechanics. Indeed, the stati of a moving object are enclosed by its position in space and the speeds of its parts: the momenta. Thus in the manifold 'positions times speeds' the possible evolutions of the object should satisfy the fact that the derivatives of the first coordinates are equal the second ones. In particular, some trajectories are not allowed. As trivial examples, you cannot vary your speed without changing your position or, similarly, you cannot move into another place at speed zero!

The 3D Heisenberg group is the most important sub-Riemannian geometry that is not in fact a Riemannian one. It is also not difficult to visualize some of its features. Topologically it is $\mathbb{R}^3$. The constrain on curves is given by what is called a 'distribution of planes'. Similarly as a smooth vector field smoothly assigns a tangent vector at each point of the manifold, a distribution of planes smoothly assigns to each point a plane (inside the 3D tangent space). The curves that will be called 'admissible' will be those curves that are tangent to such distribution.

The great feature of the Heisenberg group is that its distribution is curly enough in a way that each pair of point can be connected by at least one admissible curve. From this fact one can define a

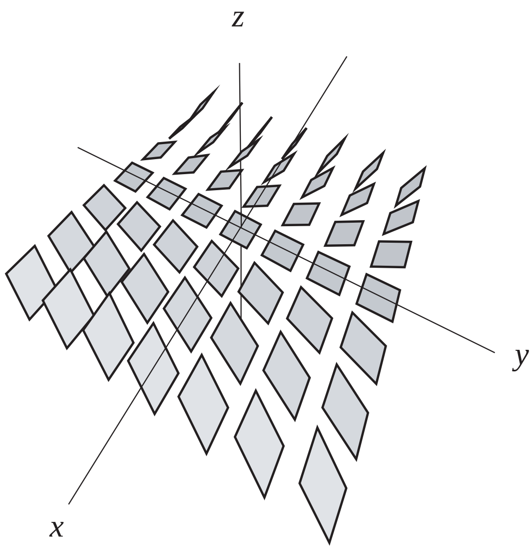


Figure 1: A standard distribution on $\mathbb{R}^3$

(finite) distance similarly to the Riemannian case: the distance between two points p and q is given by the infimum of the length of all those admissible curves from p to q,

$$d(p, q) = \inf\{\text{Length}(\gamma) \ : \ \gamma \text{ admissible, from } p \text{ to } q\}. \tag{0.0.1}$$

In the first part of this lecture notes we will discuss the following facts:

1) Such a distance d turns the space $\mathbb{R}^3$ into a metric space with the same standard topology.

2) For any two points there is in fact a geodesic curve, i.e., the distance of the points equals the length of a curve; the reader could really think that the length of such curve is its Euclidean length.

3) This metric space is really new: it is not Riemannian. It is not even biLipschitz equivalent to a Riemannian distance. In fact, the Heisenberg geometry resemble fractal geometry. Indeed, such a metric on this topologically 3-dimensional object will have metric (i.e., Hausdorff) dimension equal to 4.

The general definition of sub-Riemannian manifold follows as soon as we formalize the notion for a distribution to be 'curly enough' so that each pair of points can be connected by an admissible curve.

By a distribution on M we mean a sub-bundle of the tangent bundle TM of M. A distribution $\Delta \subseteq TM$ is called *completely non-integrable* if, for any $p \in M$, the Lie bracket algebra generated by the vector space Δ_p is the whole of T_pM. In other words, if we have that any tangent vector

$v \in T_pM$ can be presented as a linear combination of vectors of the following types

$$X_1, [X_2, X_3], [[X_4, [X_5, X_6]], \ldots,$$

where all vector fields X_j are in Δ_p.

A *sub-Riemannian manifold* is a triple (M, Δ, g), where M is a differentiable manifold, Δ is a completely non-integrable distribution and g is a smooth section of positive-definite quadratic forms on Δ. In fact, g can be considered as the restriction of a Riemannian tensor on the manifold. A curve γ on M is called admissible, or horizontal, if $\dot{\gamma}(t) \in \Delta_{\gamma(t)}$ for any t. Then the sub-Riemannian distance (also known as Carnot-Carathéodory metric) is defined by (0.0.1). Most of the previously mentioned results on the Heisenberg group will be valid for any general sub-Riemannian distance.

The understanding of many of Riemannian geometric properties come from the fact that the 'metric' tangents of a Riemannian manifold are Euclidean spaces, and the Euclidean geometry is quite well studied. Such notion of tangent is precisely defined in terms of limits of metric spaces, i.e., the tangent cones. What are the metric tangents in sub-Riemannian geometry? The answer is not immediate. For 3-dimensional sub-Riemannian manifolds we only have the Heisenberg group (another reason for it to be important). In general, alas, fixed the topological dimension, the tangent is not unique. It not the same one even for a given fixed sub-Riemannian manifold. The good news is that, analogously as the Heisenberg has a group structure, the metric tangent at 'regular' points of a sub-Riemannian manifold has a Lie group structure. It has more: it has a dilation property. Such Lie groups are those called Carnot groups.

The idea is that we should first understand the geometry of Carnot groups which are still examples of sub-Riemannian manifolds. After this, we will consider the general case of sub-Riemannian manifolds. There is hope to understand Carnot groups exactly because using the translations by elements and the dilation property it is possible to extend the theory of calculus in such a setting. The reader should notice how in the classical definition of derivative of a real function, we make use of addition, multiplications, and limits:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

All this operations are present on Carnot groups. Thus we have a metric definition of derivative, which is called today Pansu derivative, in honor to the work that Pierre Pansu did on the subject, [Pan89].

Let us enunciate one of the most celebrated theorem of Pansu, which afterwards has been expressed in its generality in [MM95].

Theorem 0.0.2 (Pansu's Rademacher Theorem [Pan89, MM95])**.** *The blow up differential of a Lipschitz map between sub-Riemannian manifolds exists at almost all points and is a group homomorphism of the tangent cones equivariant with respect to their dilations.*

In fact the theorem holds more generally for quasi-conformal mappings. The theory of quasi-conformal mappings has been used to prove rigidity theorems on hyperbolic n space over the division algebras of real, complex, or hyperbolic numbers, by studying quasi-conformal mappings on their boundary spheres at infinity.

The second part of the course will be focused on some topics of Geometric Measure Theory in the setting of Carnot groups. Most of the presented results are valid in the case of nilpotent Lie groups endowed with their Carnot-Carathéodory metric. In particular we focus on the following problems:

- Are sets that have finite perimeter rectifiable?
- How the theory of minimal surfaces differs from the Euclidean case?
- What is the regularity of geodesics?

The above questions have not complete answers yet. In fact they are leading most of the recent research in sub-Riemannian geometry.

The last part of the notes is devoted to the study of the coarse geometry of discrete nilpotent groups. We will see how a geometric notion as the polynomial growth of balls in the Cayley graph of a discrete group relates with the geometry of the tangent cone at infinity of this graph, which in this case turns out to be a Carnot group endowed with a Finsler-Carnot-Carathéodory metric, and eventually gives an algebraic consequence: the group is (virtually) nilpotent.

Chapter 1

Sub-Riemannian geometries as models

Sub-Riemannian geometry (also known as Carnot geometry in France, and non-holonomic Riemannian geometry in Russia) has been a full research domain from the 80's, with motivations and ramifications in several parts of pure and applied mathematics, namely:

* control theory

* classical mechanics

* symplectic and contact geometry

* Riemannian geometry (of which sub-Riemannian geometry constitutes a natural generalization, and where sub-Riemannian metrics may appear as limit cases)

* diffusion on manifolds

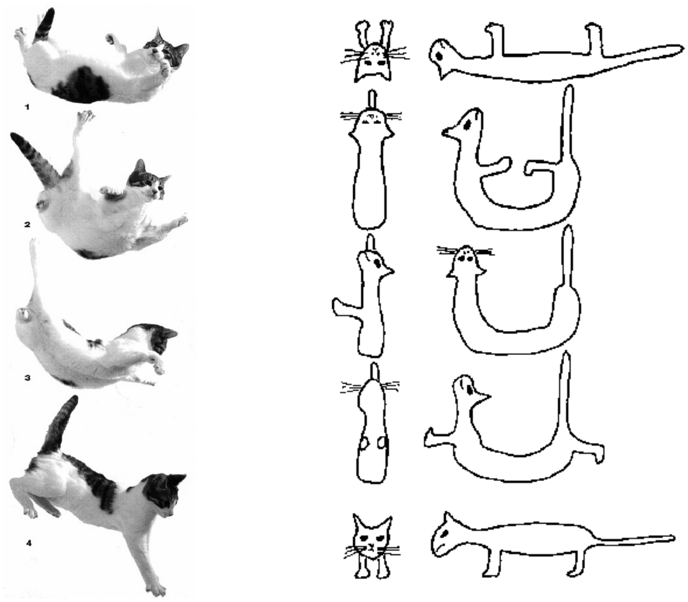
* analysis of hypoelliptic operators

* Cauchy-Riemann (or CR) geometry.

However, historically it was not clear that such theories were heading into the same notion. Thus each source provided its own jargon to the field. The non-expert reader will soon realize that some concepts coincide: a distribution of hyper-plane in an odd-dimensional manifold is in fact a contact structure and the concept of Carnot-Carathéodory metric is the same of sub-Riemannian distance.

1.1 Many examples from physics

Sub-Riemannian geometry models various structures, from finance to mechanics, from bio-medicine to quantum phases, from robots to falling cats! We don't want to enter in the details first because of lack of time, second because of lack of competence. We will address the interested reader to other



(a) A photo. (b) A sketch.

Figure 1.1: The cat spins itself around and right itself.

papers.

The geometry of principal bundles with connections

Theoretical physics defines most mechanical systems by a kinetic energy and a potential energy. Gauge theory also know as the geometry of principal bundles with connections studies systems with physical symmetries, i.e., when there is a group acting on the configuration space by isometries. Most of the times it will be easier to understand the dynamics up to isometries, successively one has to study the 'lift' of the dynamics into the initial configuration space. Such lifts will be subject to a sub-Riemannian restriction.

Falling cats

The formalism of principal bundles with connections is well presented by the example of the fall of a cat. A cat, dropped from upside down, will land on its self. The reason of this ability is the good flexibility of the cat in changing its shape.

Let us fix some formalism. Let M be the set of all the possible configurations in the 3D space of a given cat. Let S be the set of all the shapes that a cat can assume. Both M and S are manifolds of dimension quite huge. A position of a cat is just its shape plus its orientation in space. Otherwise said, the group of isometries $G := \mathrm{Isom}(\mathbb{R}^3)$ of the Euclidean 3D space acts on M and the shape

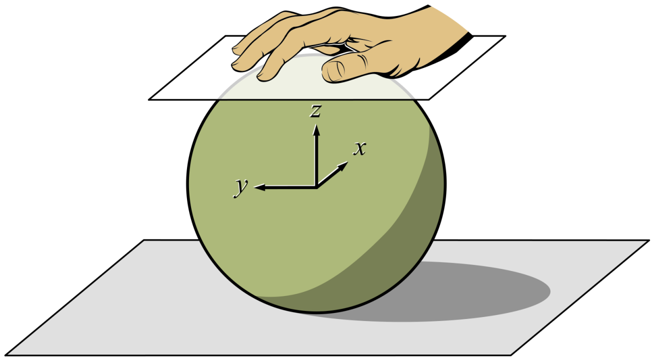


Figure 1.2: A ball rolling on the plane without sliding.

space is just the quotient of the action:

$$\pi : M \to M/G = S.$$

In fancy words, M is a principal G-bundle.

The key fact is that the cat has complete freedom in deciding its shape $\sigma(t) \in S$ at each time t. However, during the fall, each strategy $\sigma(t)$ of changing shapes will give as a result a change in configurations $\tilde{\sigma}(t) \in M$. The curve $\tilde{\sigma}(t)$ satisfies

$$\pi(\tilde{\sigma}) = \sigma.$$

Moreover the lifted curve is unique: it has to satisfy the constrain given by the 'natural mechanical connection'. What the cat is proving is that such connection has non-trivial holonomy. In other words, the cat can choose to vary its shape from the standard normal shape into the same shape giving as a result a change in configuration: the legs were initially toward the sky, then they are toward the floor.

From mechanics: parking cars, rolling balls, moving robots, and satellites

Parking a car or riding a bike. The configuration space is 3-dimensional: the position in the 2-dimensional street plus the angle with respect to a fixed line. However, the driver has only two degree of freedom: turning and pushing. Using again non-trivial holonomy we can move the car to any position we like.

Rolling a ball on the plane. A position of a ball lying on a plane requires five coordinates: two reals to characterize the point in the plane where the ball is touching it, another two coordinates to characterize the point of the ball which touches the plane, and the last one for spinning the ball around its vertical axis. When one rolls the ball without sliding, there are only three admissible control directions: two to choose a direction and then roll the ball and the third one for spinning it. Still, one can get to any position regardless of the initial position.

In *robotics* the mechanisms, as for example the arm of a robot, are subjected to constrain of movements but do not decrease the manifold of positions. Similar is the situation of *satellites*. One should really think about a satellite as a falling cat: it should choose properly its strategy of modifying the shape to have the necessary change in configuration. Another similar example is the case of an astronaut in outer space.

Vision

I became aware of the following application from conversations with S. Pauls and G. Citti. A suggested-to-curious-readers paper is [SCP08].

Neuro-biologic research over the past few decades has greatly clarified the functional mechanisms of the first layer (V1) of the visual cortex. Such layer contains a variety of types of cells, including the so-called simple cells. Researchers found that simple cells are sensitive to orientation specific brightness gradients.

Recently, this structure of the cortex has been modeled using a sub-Riemannian manifold. The space is $\mathbb{R}^2 \times \mathbb{S}^1$ where each point (x, y, θ) represents a column of cells associated to a point of retinal data $(x, y) \in \mathbb{R}^2$, all of which are attuned to the orientation given by the angle $\theta \in \mathbb{S}^1$. In other words, the vector $(\cos\theta, \sin\theta)$ is the direction of maximal rate of change of brightness at point (x, y) of the picture seen by the eye, such vector can be seen as the normal to the boundary of the picture.

The moral is that when the cortex cells are stimulated by an image, the border of the image gives a curve inside this 3D space. Such curves are restricted to be tangent to the distribution spanned by the vector fields

$$X_1 = \cos(\theta)\partial_x + \sin(\theta)\partial_y \quad \text{and} \quad X_2 = \partial_\theta.$$

Researchers think that, if a piece of the contour of a picture is missing to the eye vision (or maybe it is covered by an object), then the brain tends to 'complete' the curve by minimizing some kind of energy, in other words, there is some sub-Riemannian structure on the space of visual cells and the brain consider a sub-Riemannian geodesic between the endpoints of the missing data.

Quantum mechanical systems

I became aware of the following application from a discussion with Ugo Boscain and reading his 'Habilitation à diriger des recherches'.

Let $\mathcal{H}$ be a complex separable Hilbert space. Let us denote by $\mathbb{S}$ the unit sphere in $\mathcal{H}$.

The time evolution of quantum mechanical system (e.g., an atom, a molecule, or a system of particles with spin) is described by a map $\psi : \mathbb{R} \to \mathbb{S}$, called wave function. The vector $\psi(t)$ is called the state of the system at time t.

The equation of evolution of the state is the so-called Schrödinger equation. If the system is isolated, the equation has the form:

$$i\frac{d\psi}{dt}(t) = H_0\psi(t),$$

where H_0 is a self-adjoint operator acting on $\mathcal{H}$ called free Hamiltonian.

Let us assume for simplicity of notation that the spectrum of H_0 is discrete and non-degenerate, with eigenvalues $E_1, E_2, \ldots$ (called energy-levels) and eigenvectors $\psi_1, \psi_2, \ldots \in \mathbb{S}$.

Assume now to act on the system with some external fields (e.g an electromagnetic field) whose amplitude is represented by some functions $u_1, \ldots, u_m \in L^\infty(\mathbb{R}, \mathbb{R})$. In this case the Schrödinger equation becomes

$$i\frac{d\psi}{dt}(t) = H(t)\psi(t), \qquad \text{where } H(t) = H_0 + \sum_{j=1}^{m} u_j(t)H_j,$$

and H_j are self-adjoint operators representing the coupling between the system and the external fields. The time dependent operators $H(t)$ and $\sum_{j=1}^m u_j(t)H_j$ are called respectively the Hamiltonian and the control-Hamiltonian. The typical problem of quantum control is the so called population transfer problem:

Assume that at time zero the system is in an eigenstate ϕ_j of the free Hamiltonian H_0. Design controls $u_1, \ldots, u_m$ such that at a fixed time T the system is in another prescribed eigenstate ϕ_l of H_0.

Nowadays quantum control has many applications in chemical physics, in nuclear magnetic resonance (also in medicine) and it is central in the implementation of the so-called quantum gates (the basic blocks of a quantum computer).

For a finite dimensional quantum mechanical system, if n is the number of energy levels we have $\mathcal{H} = \mathbb{C}^n$ and the state space $\mathbb{S}$ is the unit sphere $\mathbb{S}^{2n-1} \subset \mathbb{C}^n$. In this setting, problems of quantum mechanics (being multilinear) can be formulated with matrices. The solution is of the form

$$\psi(t) = g(t)\psi(0), \qquad \text{with } g(t) \in SU(n).$$

The Schrödinger equation becomes $\frac{d}{dt}g(t) = -iH(t)g(t)$, and now $-iH(t)$ is a skew trace-zero Hermitian matrix, i.e., belongs to the Lie algebra $\mathfrak{su}(n)$.

The controllability problem (i.e. proving that for every couple of points in $SU(n)$ one can find controls steering the system from one point to the other) is nowadays well understood. Indeed, the system is controllable if and only if the Hörmander's condition holds:

$$\text{Lie}\{iH_0, iH_1, \ldots, iH_m\} = \mathfrak{su}(n).$$

Once that controllability is proved one would like to steer the system, between two fixed points in the state space, in the most efficient way. Typical costs that are interesting to minimize in the applications are:

- Energy transferred by the controls to the system (minimizing time with unbounded controls is today well understood);

- Time of transfer (minimizing time with bounded controls or energy is very difficult in general).

Even more examples

In finance... I don't know how! Talk with ETH professor Josef Teichmann.

Quantum Berry's phases... I don't know how! See references in the introduction in [Mon02].

1.2 An application to univalent function theory

There is a very remarkable application of sub-Riemmanian geometry to univalent function theory. The application is very recent and so not still well known, it is why we prefered to expose this instead of other beautiful application of sub-Riemmanian geometry to another branch of pure mathematics.

Our quick summary is based on the work by Markina, Prokhorov and Vasilev in a 2007 paper in J Funct Anal ("Sub-Riemannian geometry of the coefficients of univalent functions") and on kind conversations with Jeremy Tyson.

Classical univalent function theory considers the class S of analytic univalent functions f defined in the unit disc normalized by $f(0) = 0$ and $f'(0) = 1$.

Basic (unsolved) problems are to describe the coefficient body

$$M = \{(a_k) : (a_k) \text{ are the power series coefficients at } z = 0 \text{ for a function in } S\}$$

or its finite-dimensional slices

$$M_n = \{(a_2, a_3, \ldots, a_{n+1}) : (a_k) \text{ are the first } n \text{ (undetermined) power series coefficients at } z = 0 \text{ for a function in}$$

The Bieberbach Conjecture (proved by de Branges in 1984) says $|a_n| \leq n$ for all n. This gives information on the size of M_n and M. There is no explicit description of M_n except for the cases $n = 2$ (trivial) and $n = 3$ (Schaeffer-Spencer, 1950).

One of the basic tools in the subject is the Loewner (or Loewner-Kufarev) parametric representation, which embeds any function $f \in S$ into an ODE flow within the class S. Loewner parametrizations were used by de Branges in his proof. Nowadays there is a stochastic version of the Loewner flow (SLE) which is a very hot topic at the intersection of probability, complex analysis, stochastic PDE, math physics, etc.

Anyways, what Markina-Prokhorov-Vasilev show is that one can use the Loewner flow on S to define a natural (partially integrable) Hamiltonian system on the coefficient bodies M_n. They find certain first integrals of the flow and calculate all the relevant commutators. From there they construct a *complex* sub-Riemannian structure on M_n which is naturally adapted to the underlying univalent function theory. In fact, the Loewner parametrices become horizontal curves with restect to this sub-Riemannian structure.

Problem 1.2.1. *extending Markina-Prokhorov-Vasilev's setup to cover SLE as well as the classical (deterministic) Loewner equation.*

1.3 An isoperimetric problem on the plane

The isoperimetric problem is the problem in which, giving a length, one has to look for the maximal area among those domains with that fixed length as perimeter. We will be interested in a variant of the standard isoperimetric problem: the Dido's problem.

Dido was, according to ancient Greek and Roman sources, the founder and first Queen of Carthage (in modern-day Tunisia). She is best known from the account given by the Roman poet Virgil in his Aeneid. Indeed, in this epic poem it is narrated that King Jarbas was persuaded by Dido to give a piece of land on the African coast to settle. This land would have been as much as

Queen Dido would have encaptured with a leather string, using also the coastline. The solution is easy to find: a half-circle.

Let us give a mathematical model of such problem. On $\mathbb{R}^2$ the area form is $\mathrm{vol} = dx \wedge dy$, which is the differential of the one-form

$$\alpha = \frac{1}{2}(xdy - ydx) = \frac{1}{2}r^2 d\theta.$$

Applying Stoke's Theorem we get that, if a closed smooth counterclockwise-oriented curve γ in $\mathbb{R}^2$ encloses a domain D_γ, then the area of D_γ is just the integral of α along γ:

$$\mathrm{Area}(D_\gamma) = \iint_{D_\gamma} \mathrm{vol} = \int_\gamma \alpha.$$

Observe that at each point $(x, y) \in \mathbb{R}^2$, the vector (x, y) is in the kernel of α, thus, if L is a line through the origin, we have that $\int_L \alpha = 0$. This observation let us conclude that, if γ is a smooth curve that is not necessarily closed, then $\int_\gamma \alpha$ expresses the signed area enclosed by γ and the segment connecting the endpoints of γ.

Therefore, Dido's problem rephrases as the problem of maximize the integral $\int_\gamma \alpha$ having fixed the integral $\int_\gamma ds$, which express the length of the curve as integration of it with respect to the element of arch length ds.

1.4 The Heisenberg geometric problem

One of the models of the Heisenberg geometry is constructed as follows and it has the property that the projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ on the first two coordinates sends geodesics into those solutions of the Dido's isoperimetric problem.

If we start from a curve $\sigma(t) = (x(t), y(t))$ in $\mathbb{R}^2$, with $x(0) = y(0) = 0$, we can lift it into a 3D curve where the third coordinate $z(t)$ is the signed area encaptured by the arc $\sigma_{[0,t]}$ and the segment from 0 to $(x(t), y(t))$, see Figure 1.4. Therefore

$$z(t) := \int_{\sigma_{[0,t]}} \alpha = \int_{\sigma_{[0,t]}} \frac{1}{2}(xdy - ydx). \tag{1.4.1}$$

Differentiating in t we get

$$\dot{z} = \frac{1}{2}(x\dot{y} - y\dot{x}). \tag{1.4.2}$$

Set $\xi = dz - \frac{1}{2}(xdy - ydx)$. Consider a curve $\gamma = (\gamma_1, \gamma_2, \gamma_3) : [0, 1] \to \mathbb{R}^3$ starting at 0. Then we have that such lifted curves are exactly those satisfying $\dot{\gamma} \in \ker(\xi)$, i.e., $\xi((\dot{\gamma}_1, \dot{\gamma}_2, \dot{\gamma}_3)) \equiv 0$.

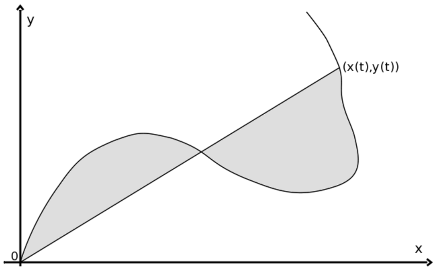
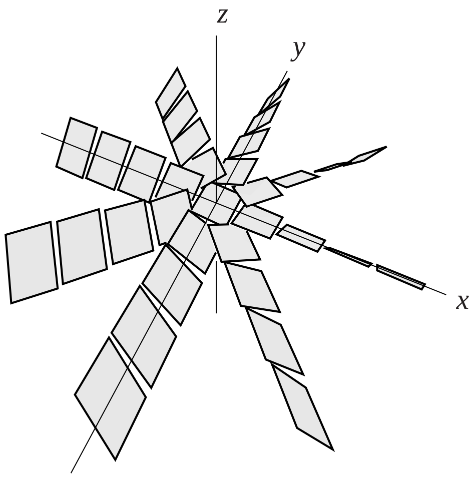
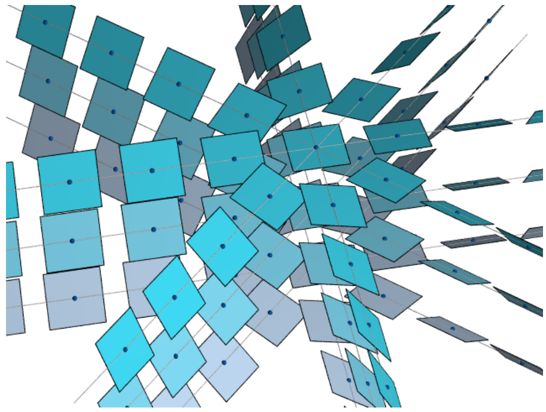


Figure 1.3: The lift of the curve is performed defining the third coordinate $z(t)$ as the oriented area of the region between the arc of the curve up to the point $(x(t), y(t))$ and the straight segment from 0 to $(x(t), y(t))$.



(a) Picture by Stephan Schönenberger.



(b) Picture by Wikipedia community.

Figure 1.4: Standard contact distribution on $\mathbb{R}^3$.

1.4.1 The standard contact structure

The previous form

$$\xi = dz - \frac{1}{2}(xdy - ydx) = dz - \frac{1}{2}r^2 d\theta \tag{1.4.3}$$

is called the 'standard contact' form[1]. It is a differential one-form on $\mathbb{R}^3$. Such form then gives at any point $(x, y, z) \in \mathbb{R}^3$ a 2D kernel inside the tangent space $T_{(x,y,z)}\mathbb{R}^3 \cong \mathbb{R}^3$ at (x, y, z):

$$\Delta_{(x,y,z)} := \ker(\xi_{(x,y,z)}) = \left\{ (v_1, v_2, v_3) \in \mathbb{R}^3 \ : \ v_3 = \frac{1}{2}(xv_2 - yv_1) \right\}.$$

[1]A contact form on a $(2n+1)$-dimensional differentiable manifold M is a 1-form α, with the property that

$$\alpha \wedge (d\alpha)^n \neq 0,$$

with

$$(d\alpha)^n = \underbrace{d\alpha \wedge \cdots \wedge d\alpha}_{n}.$$

Sometime the contact forms $dz - xdy + ydx = dz - r^2 d\theta$ and $dz + xdy$ are also called standard.

Geometrically, Δ is a field of 2D planes in the 3D space, also know as distribution. Now, given vectors $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$, consider the linear product given by

$$\langle v, w\rangle := v_1 w_1 + v_2 w_2. \tag{1.4.4}$$

Notice that, since the planes $\Delta_{(x,y,z)}$ never in the z-axis, then the restriction of $\langle \cdot, \cdot \rangle$ on $\Delta_{(x,y,z)}$ is a positive-defined inner product. If one prefers, such restriction could be thought as a restriction of a Riemannian tensor on $\mathbb{R}^3$, i.e., a positive-defined inner product on the whole of the tangent bundle of $\mathbb{R}^3$. Indeed, we can fix the following frame[2] of $\mathbb{R}^3$:

$$\begin{cases} X &= \frac{\partial}{\partial x} - \frac{1}{2} y \frac{\partial}{\partial z}, \\ Y &= \frac{\partial}{\partial y} + \frac{1}{2} x \frac{\partial}{\partial z}, \\ Z &= \frac{\partial}{\partial z}, \end{cases} \tag{1.4.5}$$

and declare it orthonormal. Let us check that such Riemannian metric gives the linear product (1.4.4) when restricted to the plane $\Delta_{(x,y,z)}$. Since $\frac{\partial}{\partial x} = X + \frac{1}{2} yZ$ and $\frac{\partial}{\partial y} = Y - \frac{1}{2} xZ$, then

$$v = v_1 X + v_2 Y + (\frac{v_1}{2} y - \frac{v_2}{2} x + v_3) Z.$$

So, if $v \in \Delta_{(x,y,z)}$, we have $v = v_1 X + v_2 Y$ and thus (1.4.4).

In contact geometry a curve γ is called Legendrian with respect to ξ if $\xi(\dot\gamma) \equiv 0$. In other words, if the tangent vector $\dot\gamma(t)$ lies in the plane $\Delta_{\gamma(t)}$. Given a Legendrian curve γ, we define its length $L(\gamma)$ as the integral of the norm of $\dot\gamma$ with respect to the scalar product (1.4.4). In other words, $L(\gamma)$ is exactly the Euclidean length of the projection of γ onto the first two components of $\mathbb{R}^3$.

At this point we introduce a new distance on $\mathbb{R}^3$ which we refer to it as the contact distance. For any p and q in $\mathbb{R}^3$, define

$$d_c(p, q) := \inf\{L(\gamma) \ : \ \gamma \text{ Legendrian between } p \text{ and } q\}. \tag{1.4.6}$$

The fact that ξ was obtained from the Dido's problem tells us that for any pair of points there are several Legendrian curves joining it:

A crucial fact: Any pair of points is connected by a curve that is Legendrian with respect to ξ. Indeed, to connect say $(0, 0, 0)$ to (x, y, z), it is enough to take a curve σ on $\mathbb{R}^2$ from $(0, 0)$ to (x, y) with the property that the signed area enclosed by σ and the segment from $(0, 0)$ to (x, y) is exactly z. Then the lifted curve $\tilde\sigma$ will connected $(0, 0, 0)$ to (x, y, z).

Moreover we also know that the Riemannian length of $\tilde\sigma$ equals the planar Euclidean length of σ. Therefore, there is a correspondence between geodesics with respect to the metric d_c and solutions

[2] A frame is a set of vector fields on a differentiable manifold M that at each point $p \in M$ gives a basis of T_pM.

of the 'dual' Dido's isoperimetric problem: fixed a value for the area, minimize the perimeter. Since it is easy to find solutions of Dido's problem we will be able to write explicitly the geodesics of the metric space $(\mathbb{R}^3, d_c)$. We will do this later in Section 2.5.

1.5 The Heisenberg group

A crucial property of the Heisenberg geometry is that the space is isometrically homogeneous. In fact, $\mathbb{R}^3$ can be endowed with a group structure (different from the Euclidean one) in such a way that all of the above constructions are preserved by the action of the group onto itself.

Indeed, consider the group law[3]

$$(x, y, z) \cdot (x', y', z') := \left(x + x', y + y', z + z' + \frac{1}{2}(xy' - yx') \right). \tag{1.5.1}$$

We claim that the left translations preserve the distribution Δ and moreover preserve the orthogonal frame X, Y, Z defined by (1.4.5). Let's verify this claim for X. Call f a fixed left translation

$$f(x, y, z) := L_{(s,t,u)}(x, y, z) = (s, t, u) \cdot (x, y, z) = \left(x + s, y + t, z + u + \frac{1}{2}(sy - tx) \right). \tag{1.5.2}$$

The differential is

$$df = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -t/2 & s/2 & 1 \end{pmatrix}. \tag{1.5.3}$$

So $df X = \frac{\partial}{\partial x} + \left(-\frac{t}{2} - \frac{y}{2}\right) \frac{\partial}{\partial z}$. On the other hand, $X \circ f = \frac{\partial}{\partial x} - \frac{1}{2}(t + y) \frac{\partial}{\partial z}$. Therefore $f_* X = X \circ f$, i.e., X is left-invariant. Analogously, $f_* Y = \frac{\partial}{\partial y} + \frac{1}{2}(s + x) \frac{\partial}{\partial z} = Y \circ f$ and $f_* Z = \frac{\partial}{\partial z} = Z \circ f$.

The next proposition summarizes the above discussion.

Proposition 1.5.4. *The Heisenberg geometry is isometrically homogeneous: the space has a Lie group structure so that each left translation is an isometry with respect to the contact distance d_c.*

The above model of the Heisenberg group has the advantage that it is easy to compute and visualize its 1-dimensional subgroups. Indeed, one-parameter subgroups for this group structure are the standard Euclidean lines:

$$\gamma_v(t) = \exp\left(t(v_1, v_2, v_3)\right) = (tv_1, tv_2, tv_3).$$

In particular, all lines through 0 in the xy-plane are geodesics.

[3]The reader can check that the above one is a group structure and it turns $\mathbb{R}^3$ into a Lie group, i.e., multiplication and inversion are smooth maps.

1.5.1 The 3D nilpotent (non-Abelian) matrix group

The Heisenberg group has also a matrix model. It can be seen as a subgroup of the group of invertible matrices. The Heisenberg group is the group of 3×3 upper triangular matrices equipped with the usual matrix product:

$$\mathbb{G} = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \ : \ a, b, c \in \mathbb{R} \right\} < GL(3, \mathbb{R}).$$

Such a model is useful because (first, it is easy to remember the group structure! then) the Lie algebra can be also seen as a matrix group and the exponential of the Lie group is the classical exponential of matrices. Indeed, the Lie algebra is

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \ : \ a, b, c \in \mathbb{R} \right\}.$$

A basis of the Lie algebra is

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.5.5}$$

One parameter subgroups are of the form:

$$\begin{aligned} \gamma_{(a,b,c)}(t) &= \exp \left(t \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \right) \\ &= I + t \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} + t^2 \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}^2 + \dots \\ &= I + t \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} + t^2 \begin{pmatrix} 0 & 0 & ab \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 0 \\ &= \begin{pmatrix} 1 & at & ct + abt^2 \\ 0 & 1 & bt \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

1.5.2 The uniqueness of the Heisenberg group

We claim that the map

$$\varphi : (x, y, z) \mapsto \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

is a Lie group isomorphism from the Lie group $\mathbb{R}^3$ with the product (1.5.1) to the Lie group of 3×3 upper triangular matrices with the usual matrix product. Indeed, the map φ is a group homomorphism (straightforward calculation) and its differential at the identity is the identity matrix.

In fact, more is true. The Heisenberg group is a nilpotent group and there are only two 3D simply-connected nilpotent Lie groups: the Euclidean 3-space and the Heisenberg group.

Indeed,consider the Lie algebra $\mathfrak{g}$ of the group. Since $\mathfrak{g}$ is nilpotent, one can take Z in the center of $\mathfrak{g}$ which is non-trivial.Complete Z to a basis X, Y, Z of $\mathfrak{g}$. Now, either X and Y commute, and so the algebra is commutative, or $W := [X, Y] \neq 0$. write $W = aX + bY + cZ$. Then $[W, Y] = aW$ and so, since $\mathfrak{g}$ is nilpotent, we have $a = 0$. Analogously $b = 0$. Thus $c \neq 0$, and, replacing Z with cZ, we have that the algebra of $\mathfrak{g}$ is defined by the relations:

$$[X, Y] = Z \quad \text{and} \quad [X, Z] = [Y, Z] = 0.$$

We can conclude the proof recalling that there exists a unique simply-connected Lie group with a fixed Lie algebra (see Section 3.1.7)

1.6 Exercises

1. Prove dido's solution: the maximal area enclosed by a curve of length l on the plane together with a fixed line is $2l^2/\pi$ and it is only obtained as an half disk.

2. Let $\mathrm{vol} = dx \wedge dy$ and $\alpha = \dfrac{1}{2}(xdy - ydx)$. Prove

   (a) $d(\mathrm{vol}) = \alpha$;

   (b) in polar coordinate, we have $\alpha = \dfrac{1}{2} r d\theta$;

   (c) if L is a line through the origin, then $\int_L \alpha = 0$.

3. Let σ be a Lipschitz curve on the plane. Let $\sigma_{[0,t]} = (x(t), y(t))$ be the arc up to time t. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. Show that

$$\frac{d}{dt}\left(\int_{\sigma_{[0,t]}} f(x, y)dx\right) = f(x(t), y(t))\frac{dx}{dt}(t), \qquad \text{almost everywhere.}$$

4. Show the relations

$$[X, Y] = Z \quad \text{and} \quad [X, Z] = [Y, Z] = 0.$$

   in the following cases:

   (a) for the vector fields in (1.4.5),

   (b) for the matrices (1.5.5).

5. Calculate the inverse of an element (x, y, z) with respect to the group structure given by (1.5.1).

6. Consider the group structure on $\mathbb{R}^3$ given by (1.5.1). Prove that the lines

$$\gamma_v(t) = (tv_1, tv_2, tv_3)\,.$$

are one-parameter subgroups.

7. Let L be a line through 0 in the xy-plane of $\mathbb{R}^3$. Prove that L is a geodesic with respect to the contact distance distance d_c defined in (1.4.6).

8. Consider the map

$$\varphi : (x, y, z) \mapsto \begin{pmatrix} 1 & x & z + \frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

from $\mathbb{R}^3$ with the product (1.5.1) to the space of 3×3 upper triangular matrices with the usual matrix product. Prove that

(a) the map is a Lie group isomorphism,

(b) the map send the standard basis X, Y, and Z (defined in (1.4.5)) of the first Lie algebra to the standard basis X, Y, and Z (defined in (1.5.5)) of the second Lie algebra.

9. On the vertical z-axis the distance d_c defined in (1.4.6) is a multiple of the square root of the Euclidean one.

Chapter 2

The general theory

2.1 A review of metric geometry: lengths, geodesics spaces, and Hausdorff measures

An overview of the main notions is necessary to clarify the setting and the terminology. There are several excellent books [Fed69, Gro99, AFP00, Hei01, BBI01] giving a clear and detailed exposition of the material. The purpose here is to comment some facts for non-experts.

A metric space will be denoted as (X, d), where X is the set and d the distance. If it is clear what metric we are considering, we could write just X for the metric space. We will use the term metric as a synonym of distance function, and never as a shortening of 'Riemannian metric'.

A *curve* or *path* γ is a continuous mapping $\gamma : I \to X$, where $I \subset \mathbb{R}$ is an interval and X is a topological space. The map γ will often be conflated with its image $\gamma(I)$.

In a metric space (X, d), the *length of a curve* $\gamma : [a, b] \to X$ is

$$L(\gamma) = \mathrm{Length}_d(\gamma) := \sup \left\{ \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1})) : n \in \mathbb{N} \text{ and } a = t_0 < t_1 < \cdots < t_n = b \right\}.$$

A *rectifiable curve* is a curve with finite length. One might easily check that the length does not depend on the parametrization. A parametrization of γ is called of unit speed or parametrized by arc length if for any t_1, t_2 in $[a, b]$, we have

$$\mathrm{Length}(\gamma|_{[t_1,t_2]}) = |t_2 - t_1|.$$

Whenever the distance $d(x, y)$ is equal to the infimum of the lengths of all paths from x to y we say that (X, d) is a *length space* or a *path metric space* or that the metric is *intrinsic*. If moreover the infimum is attained then the space is called *geodesic*. In other words a geodesic space is a metric space where any two points are the endpoints of a curve whose length is exactly the distance between the two points. Such a curve is called a *geodesic*. Not all length spaces are geodesic spaces, one reason can be lackness of completeness. For locally compact spaces this can be the only obstruction:

Theorem 2.1.1 (Hopf-Rinow-Cohn-Vossen [BBI01, Theorem 2.5.23])**.** *If a length space (X, d) is complete and locally compact then any two points in X can be connected by a geodesic.*

Given two metric spaces (X, d_X) and (Y, d_Y), a function

$$f : X \to Y$$

is called *Lipschitz* if there exists a real constant $K \geq 0$ such that, for all x_1 and x_2 in X,

$$d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2).$$

The value K (or many times the smallest value of such K's) is called the *Lipschitz constant* of the function f. A function is called *locally Lipschitz* if for every $x \in X$ there exists a neighborhood U of x such that f restricted to U is Lipschitz.

If there exists a $K \geq 1$ with

$$\frac{1}{K} d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2),$$

then f is called *biLipschitz* (also written bi-Lipschitz or bilipschitz). The biLipschitz homeomorphisms are the isomorphisms in the category of Lipschitz maps. To be more explicit on the value of the constant K we would say that f is a K-biLipschitz map. A biLipschitz mapping is injective, and is in fact a homeomorphism onto its image. A 1-biLipschitz map is called an *isometry*.

Let S be any subset of X, and $\delta > 0$ a real number. Define

$$H^d_\delta(S) := \inf\Big\{\sum_{i=1}^{\infty} \mathrm{diam}(U_i)^d : \bigcup_{i=1}^{\infty} U_i \supset S, \, \mathrm{diam}(U_i) < \delta\Big\}.$$

Then

$$\mathcal{H}^d(S) := \sup_{\delta > 0} H^d_\delta(S) = \lim_{\delta \to 0} H^d_\delta(S)$$

is called the d-dimensional *Hausdorff measure* of S. The *Hausdorff dimension* is

$$\dim_{\mathrm{Haus}}(S) := \inf\{d \geq 0 : \mathcal{H}^d(S) = 0\} = \sup\big(\{d \geq 0 : \mathcal{H}^d(S) = \infty\} \cup \{0\}\big).$$

2.2 A review of differential geometry

2.2.1 Vector fields and Lie brackets

We will denote by M a smooth differentiable manifold with topological dimension n.

For $x \in M$, the fiber T_xM of the tangent bundle TM is a derivation of germs of C^∞ functions at x (i.e., an $\mathbb{R}$-linear application from $C^\infty(x) \to \mathbb{R}$ that satisfies the Leibnitz rule). If $F : M \to N$ is smooth and $x \in M$, we shall denote by $dF_x : T_xM \to T_{F(x)}N$ its differential, defined as follows: the pull back operator $u \mapsto F_x^*(u) := u \circ F$ maps $C^\infty(F(x))$ into $C^\infty(x)$; thus, for $v \in T_xM$ we have that

$$dF_x(v)(u) := v(u \circ F)(x), \qquad u \in C^\infty(F(x))$$

defines an element of $T_{F(x)}N$.

We denote by $\Gamma(TM)$ the linear space of smooth vector fields, i.e. smooth sections of the tangent bundle TM; we will typically use the notation X, Y, Z to denote them. We use the notation $[X,Y]f := X(Yf) - Y(Xf)$ for the Lie bracket, that induces on $\Gamma(TM)$ an infinite-dimensional Lie algebra structure.

If $F : M \to N$ is smooth and invertible and $X \in \Gamma(TM)$, the push forward vector field $F_*X \in \Gamma(TN)$ is defined by the identity $(F_*X)_{F(x)} = dF_x(X_x)$. Equivalently,

$$(F_*X)u := [X(u \circ F)] \circ F^{-1} \qquad \forall u \in C^\infty(M). \tag{2.2.1}$$

The push-forward commutes with the Lie bracket, namely

$$[F_*X, F_*Y] = F_*[X,Y] \qquad \forall X,\, Y \in \Gamma(TM). \tag{2.2.2}$$

If $F : M \to N$ is smooth and σ is a smooth curve on M, then

$$dF_{\sigma(t)}(\sigma'(t)) = (F \circ \sigma)'(t), \tag{2.2.3}$$

where $\sigma'(t) \in T_{\sigma(t)}M$ and $(F \circ \sigma)'(t) \in T_{F(\sigma(t))}N$ are the tangent vector fields along the two curves, in M and N. If $u \in C^\infty(M)$, identifying $T_{u(p)}\mathbb{R}$ with $\mathbb{R}$ itself, given $X \in \Gamma(TM)$, we have

$$du_p(X) = X_p(u).$$

2.2.2 Riemannian and Finsler geometry

Let M be a differentiable manifold of dimension n. A *Riemannian metric* on M is a family of (positive definite) inner products

$$g_p : T_pM \times T_pM \longrightarrow \mathbb{R}, \qquad p \in M,$$

such that, for all differentiable vector fields X, Y on M,

$$p \mapsto g_p(X_p, Y_p)$$

defines a differentiable function $M \to \mathbb{R}$. The assignment of an inner product g_p to each point p of the manifold is called a metric tensor.

In a system of local coordinates on the manifold M given by n real-valued functions $x_1, \ldots, x_n$, the vector fields

$$\left\{ \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right\}$$

give a basis of tangent vectors at each point of M. Relative to this coordinate system, the components of the metric tensor are, at each point p,

$$g_{ij}(p) := g_p\left(\left(\frac{\partial}{\partial x_i} \right)_p, \left(\frac{\partial}{\partial x_j} \right)_p \right).$$

Equivalently, the metric tensor can be written in terms of the dual basis $\{dx_1, \ldots, dx_n\}$ of the cotangent bundle as

$$g = \sum_{i,j} g_{ij} \mathrm{d}x_i \otimes \mathrm{d}x_j.$$

Endowed with this metric, the differentiable manifold (M, g) is called a Riemannian manifold.

Finsler manifolds generalize Riemannian manifolds by no longer assuming that they are infinitesimally Euclidean in the sense that the norm on each tangent space is necessarily induced by an inner product. Two good references on Finsler geometry are [BCS00] and [AP94].

A Finsler structure on a differentiable manifold M is given by a function $\|\cdot\| : TM \to \mathbb{R}$ that is smooth on the complement of the zero section of TM and such that the restriction of $\|\cdot\|$ to any tangent space T_pM is a (symmetric) norm[1]. A Riemannian manifold has an induced Finsler structure $\|v\| = \sqrt{g(v,v)}$.

Connected Riemannian and Finsler manifolds carry the structure of length metric spaces.

Specifically, let $(M, \|\cdot\|)$ be a connected Finsler manifold. Let $\gamma : [a, b] \to M$ be a parametrized curve in M, which is differentiable with velocity vector $\dot{\gamma}$. The length of γ is defined as

$$L(\gamma) := \int_a^b \|\dot{\gamma}(t)\| \, \mathrm{d}t.$$

[1] The definition of Finsler structure could be even more generalized allowing, for example, asymmetric norms or not assuming linearity. Some authors also assume that a Finsler structure has strongly convex unit spheres, we do not.

By change of variables, the arc-length is independent of the chosen parametrization. In particular, a curve $\gamma : [a,b] \to M$ can be parametrized by its arc length. A curve is parametrized by arc-length if and only if $\|\dot\gamma(t)\| = 1$, for all $t \in [a,b]$.

The distance function $d : M \times M \to [0, +\infty)$ is defined by

$$d(p,q) = \inf L(\gamma),$$

where the infimum extends over all differentiable curves γ beginning at $p \in M$ and ending at $q \in M$.

This function d satisfies the properties of a distance function for a metric space. The only property which is not completely straightforward is that $d(p,q) = 0$ implies $p = q$. For this property, if the manifold is Riemannian, one can use a normal coordinate system, which also allows one to show that the topology induced by d is the same as the original topology on M. For Finsler manifolds one realizes that any Finsler structure is locally biLipschitz equivalent to a Riemannian structure.

2.3 The definition of Finsler-Carnot-Carathéodory distances

Let $(M, \|\cdot\|)$ be a smooth Finsler manifold. Let $\Delta \subseteq TM$ be a sub-bundle[2] of the tangent bundle of M. The names *distribution* or polarization are also used to refer to such a Δ. The triple $(M, \Delta, \|\cdot\|)$ is called *sub-Finsler manifold*; if the Finsler norm $\|\cdot\|$ is given by a Riemannian scalar product g, then (M, Δ, g) is called *sub-Riemannian manifold*. An absolutely continuous[3] curve γ in M is said to be *horizontal* with respect to the distribution Δ if $\dot\gamma(t) \in \Delta$, for almost all t. The names *admissible or controlled paths* are also used to refer to such a γ. The length of γ with respect to $\|\cdot\|$ is

$$L(\gamma) := \mathrm{Length}_{\|\cdot\|}(\gamma) := \int_0^1 \|\dot\gamma(t)\|\, dt.$$

Then one can consider the metric on M induced by Δ and $\|\cdot\|$.

For any $x, y \in M$, we define the distance function

$$d_{CC}(x,y) := \inf\{\mathrm{Length}_{\|\cdot\|}(\gamma) \mid \gamma \text{ horizontal, from } x \text{ to } y\}. \tag{2.3.1}$$

Roughly speaking, in presence of minimizing curves, the distance of two points is given by the length of the shortest horizontal curve. Such a distance is called the *Finsler-Carnot-Carathéodory distance.*

[2] A subbundle E of a vector bundle F on a topological space M is a collection of linear subspaces E_p of the fibers F_p of F at p in M, that make up a vector bundle in their own right.

[3] A curve $f : [a,b] \to \mathbb{R}^n$ is absolutely continuous if and only if there is a Lebesgue integrable map $\dot f : [a,b] \to \mathbb{R}^n$ such that

$$f(x) = f(a) + \int_a^x \dot f(x) dx.$$

When the norm is given by a scalar product, the distance d_{CC} defined in (2.3.1) is called *Carnot-Carathéodory* or *sub-Riemannian.* Such metric appeared in the literature under a variety of names: 'singular Riemannian metric', 'non-holonomic Riemannian metric'. It was also used in the theory of hypo-elliptic PDE, but without a name.

Remark 2.3.2. Notice that we have assumed yet nothing that would imply that such a distance is finite. The metric d_{CC} is finite provided that the distribution Δ satisfies the following condition due to Hörmander, and the manifold M is assumed connected.

2.3.1 Hörmander's condition

To describe Hörmander's condition, let $X_1, X_2, \ldots, X_k$ be a local framing for Δ near a point $p \in M$. If these vector fields, along with all their commutators, span T_pM, then the vector fields are said to satisfy Hörmander's condition at p.

Denote by $\Delta^{[j]}(p)$ the subspace of T_pM spanned by all commutators of the X_i's of order $\leq j$ (including, of course, the X_i's). Namely, $X_i(p)$ has order 1; $[X_i, X_j](p)$ has order 2; $[X_i, [X_j, X_k]](p)$ has order 3; but those of order 4 are those in one of the two forms:

$$[X_i, [X_j, [X_k, X_l]]](p) \qquad \text{or} \qquad [[X_i, X_j], [X_k, X_l]](p).$$

Once can show, see Exercise 2.7.5, that $\Delta^{[j]}(p)$ does not depend upon the choice of local basis X_j, so it makes sense to say that the distribution satisfies *Hörmander's condition at* p if $\dim \Delta^{[j]}(p) = \dim(M)$ for some j, i.e.,

$$\text{(Hörmander's condition)} \qquad T_pM = \Delta^{[j]}(p), \qquad \text{for some } j \in \mathbb{N}.$$

The sub-bundles $\Delta \subseteq TM$ that satisfy the *Hörmander's condition at every point* of M are also called *bracket generating distributions.*

The number

$$Q = \sum_j j \left(\dim \Delta^{[j]}(p) - \dim \Delta^{[j-1]}(p) \right) \tag{2.3.3}$$

is called the *metric dimension* or also *homogeneous dimension* (at p) and will have an important metric role, see Corollary 2.6.6.

2.3.2 The generalization of Control Theory

In Control Theory one is interested in systems of differential equations of the form

$$\dot{u} = \sum_{i=1}^{m} c_j(t) X_{(u)}, \tag{2.3.4}$$

where $X_1, \ldots, X_m$ are given vector fields on M, and the $c_1, \ldots, c_m$ are variable L^1 functions on some bounded interval. These functions are called *control functions* or *controls.* Any path obtained integrating (2.3.4) is called a controlled path.

When the rank of the system of vector fields $X_1, \ldots, X_m$ is constant, controlled paths coincide with the absolutely continuous paths tangent to the distribution

$$\Delta = \mathbb{R}\text{- span}\langle X_1, \ldots, X_m \rangle$$

generated by $X_1, \ldots, X_m$. Conversely, any rank m distribution Δ can, locally, be written as $\Delta = \langle X_1, \ldots, X_m \rangle$. Observe that in the previous sentence, the adverb 'locally' is needed, for global topological reasons, as for example for $\Delta = T(S^2)$.

However, for many systems of interest in Control Theory, the rank of $X_1, \ldots, X_m$ is not constant, but one can still define a related distance: for $p \in M$ and $v \in T_pM$, set

$$g_p(v) := \inf\{u_1^2 + \cdots + u_m^2 \mid u_1X_1 + \cdots + u_mX_m = v\}.$$

We are using the notation that $\inf \emptyset = +\infty$. We then have that g_p is a positive definite quadratic form on the subspace

$$\Delta_p := \mathbb{R}\text{- span}\langle X_1(p), \ldots, X_m(p)\rangle.$$

The *control distance associated to the system* $X_1, \ldots, X_m$ is defined as, for any p and q in M,

$$d(p,q) = \inf \left\{ \int_0^1 g_p(\dot\gamma(t))^{1/2} dt \;\middle|\; \gamma \text{ absolutely continuous path } \gamma(0) = p, \gamma(1) = q \right\}. \tag{2.3.5}$$

2.3.3 The general definition

Definition 2.3.6. *A (smooth) sub-Finsler structure on a manifold M is a function $g : TM \to [0, \infty]$ obtained by the following construction: Let E be a vector bundle over M endowed with a norm $|\cdot|$ and let*

$$\sigma : E \to TM$$

be a morphism of vector bundles. For each $p \in M$ and $v \in T_pM$, set

$$g_p(v) := \inf\{|u| : u \in E_p, \sigma(u) = v\}.$$

Analogously as before, one define the *sub-Riemannian distance associated to the bundle E*, for any p and q in M, as

$$d(p,q) = \inf \left\{ \int_0^1 g_p(\dot\gamma(t))^{1/2} dt \;\middle|\; \gamma \text{ absolutely continuous path } \gamma(0) = p, \gamma(1) = q \right\}.$$

One can check that, for the inclusion $\sigma : \Delta \hookrightarrow TM$ of a sub-bundle of the tangent bundle, one recovers the Finsler-Carnot-Carathéodor distance (2.3.1). For $E = M \times \mathbb{R}^m$ and $\sigma(p, v) := u_1 X_1 + \cdots + u_m X_m$, one recovers the control distance (2.3.5).

2.3.4 Energy VS Length

The energy of a parametrized curve $\gamma : [0,1] \to M$ with respect to $\|\cdot\|$ is

$$E(\gamma) := \mathrm{Energy}_{\|\cdot\|}(\gamma) := \sqrt{\int_0^1 \|\dot{\gamma}(t)\|^2 \, dt}.$$

One has the equality:

$$d_{CC}(x, y) := \inf\{\mathrm{Energy}_{\|\cdot\|}(\gamma) \mid \gamma \text{ horizontal, from } x \text{ to } y\}. \tag{2.3.7}$$

On the contrary of length, energy depends on the parametrization of the curve. However, by CauchySchwarz inequality, we alway have

$$L(\gamma) \le E(\gamma),$$

and equality in the case that the curve is parametrized by a multiple of the arc length.

2.4 Chow's Theorem and geodesic existence

We want to motivate now the fact that if a sub-Riemannian manifold has the property that its distribution satisfies the Hörmander's condition, then the Carnot-Carathéodory distance is finite.

Hörmander's condition can be considered as an infinitesimal transitivity. Chow's Theorem implies local transitivity:

Theorem 2.4.1 (Chow)**.** *If a smooth distribution satisfies Hörmander's condition at some point p, then any point q which is sufficiently close to p can be joined to p by a horizontal curve.*

A consequence of the proof of the theorem is that close points can be connected by short curves.

Theorem 2.4.2 (Chow)**.** *If M is connected and Hörmander's condition holds at any point, then the Carnot-Carathéodory distance d_{CC} is finite. Moreover, the topology defined by d_{CC} is the original topology of M.*

A sketchy proof of Theorem 2.4.1. We present here a proof taking for grant a theorem by Sussmann. We are omitting the proof of Sussmann's Theorem which is in fact the core of Chow's Theorem, but it is well presented in [Bel96]. Later in the notes we will give a detailed proof in the easier case of Carnot groups.

Theorem 2.4.3 (Sussmann [Sus73, Ste74, Bel96]). *The sets of points accessible from a given point p in M is an immersed sub-manifold.*

Call $\Sigma \subset M$ the immersed sub-manifold of points accessible from p. Let Δ be the distribution. Now, given vector field $X \in \Delta$, consider the integral curve γ solution of the equation:

$$\gamma(0) = p, \text{ and } \dot{\gamma}(t) = X_{\gamma(t)},$$

then obviously γ is horizontal. Therefore the vector X_p is tangent to Σ. Thus the whole of Δ_p is tangent to Σ. Since by Sussmann's Theorem Σ is an immersed sub-manifold, then its tangent $T_p\Sigma$ is closed under bracket. Therefore the Lie span of Δ_p is tangent to Σ. Then since Hörmander's condition holds at p, then T_pM is tangent to Σ. From this we have $\dim M = \dim \Sigma$, and thus Σ is a neighborhood of p. □

Theorem 2.4.4 (Hopf-Rinow Theorem for sub-Riemannian manifolds). *If a distribution $\Delta \subset TM$ satisfies the Hörmander's condition at any point in M. Then sufficiently near points can be joined by a geodesic with respect to d_{CC}. Moreover, if M is connected and the metric space (M, d_{CC}) is complete, then any two points in M can be connected by a geodesic.*

Proof. Theorem 2.4.2 assets that the topology of the metric space (M, d_{CC}) is the same of the manifold topology. In particular the space is locally compact. Applying Ascoli-Arzelà Theorem to a compact ball we have existence of geodesics at small scale. Applying the general Hopf-Rinow Theorem for complete and locally compact length spaces (see Theorem 2.1.1), we have the existence of global geodesics. □

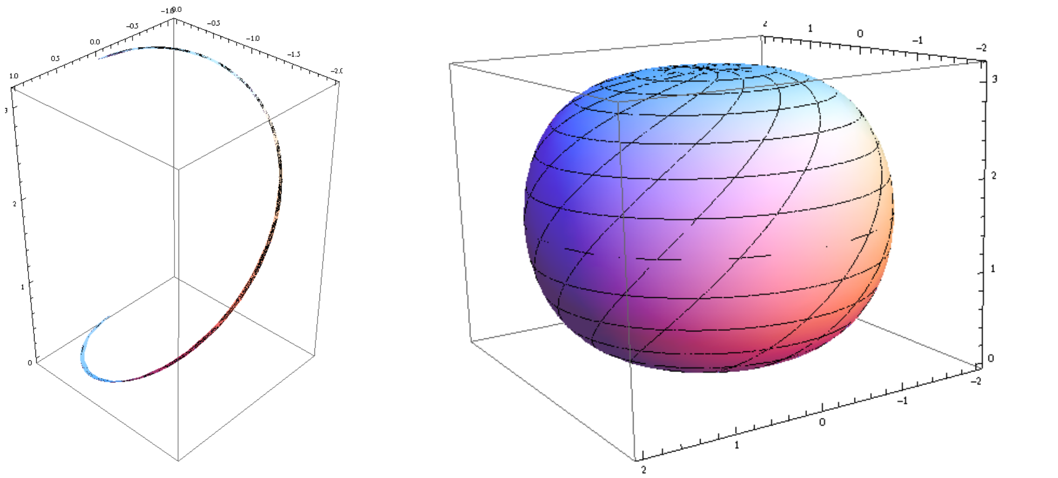
2.5 The geodesics and the spheres in the Heisenberg group

We want now to use the fact that the we know the solution of the isoperimetric problem together with the previous discussion at the end of Section 1.4.1, to have explicit formulae for the geodesics in the Heisenberg group. Briefly, we saw that for how the geometry in the Heisenberg group has been constructed, then the geodesics in this metric space starting at 0 are the lift of the shortest curves on the plane that enclose a fixed area, which are part of circles. Thus the geodesic from the identity in the Heisenberg group are the lift of circles.

The curve
$$\left(\frac{\cos(kt)-1}{k},\frac{\sin(kt)}{k}\right),\ \text{for } t\in[0,2\pi/|k|],$$
is the parametrization by arc length (starting at $(0,0)$) of the circle of perimeter $2\pi/|k|$, with center on the negative x-axis and passing through 0. For $k=0$, we can still consider the formula in the limit sense: the circles degenerate to the line $(0,t)$, defined for all $t\in\mathbb{R}$.

We may rotate the curves by an angle $\theta\in\mathbb{R}/2\pi\mathbb{Z}$:
$$\begin{aligned}
R_\theta\left(\frac{\cos(kt)-1}{k},\frac{\sin(kt)}{k}\right) &= \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\left(\frac{\cos(kt)-1}{k},\frac{\sin(kt)}{k}\right)\\
&= \left(\cos\theta\frac{\cos(kt)-1}{k}-\sin\theta\frac{\sin(kt)}{k},\sin\theta\frac{\cos(kt)-1}{k}+\cos\theta\frac{\sin(kt)}{k}\right).
\end{aligned}$$
We can calculate the third coordinate as in (1.4.1).
$$\begin{aligned}
z(T) &= \int_0^T\frac{1}{2}(xdy-ydx)\\
&= \frac{1}{2}\int_0^T\left(\cos\theta\frac{\cos(kt)-1}{k}-\sin\theta\frac{\sin(kt)}{k}\right)(-\sin\theta\sin(kt)+\cos\theta\cos(kt))+\\
&\qquad-\left(\sin\theta\frac{\cos(kt)-1}{k}+\cos\theta\frac{\sin(kt)}{k}\right)(-\cos\theta\sin(kt)-\sin\theta\cos(kt))\,dt\\
&= \frac{1}{2k}\int_0^T-\cos\theta(\cos(kt)-1)\sin\theta\sin(kt)+(\cos\theta)^2(\cos(kt)-1)\cos(kt)+\\
&\qquad+(\sin\theta)^2\left(\sin(kt)\right)^2-\sin\theta\sin(kt)\cos\theta\cos(kt)+\\
&\qquad+\sin\theta(\cos(kt)-1)\cos\theta\sin(kt)+(\sin\theta)^2(\cos(kt)-1)\cos(kt)+\\
&\qquad+(\cos\theta)^2\left(\sin(kt)\right)^2+\cos\theta\sin(kt)\sin\theta\cos(kt)dt\\
&= \frac{1}{2k}\int_0^T(\cos(kt)-1)\cos(kt)+(\sin(kt))^2\,dt\\
&= \frac{1}{2k}\int_0^T 1-\cos(kt)dt=\frac{1}{2k^2}(Tk+\sin(kT)).
\end{aligned}$$

We conclude that geodesics starting from the origin $0\in\mathbb{R}^3$ are smooth curves $\gamma=(\gamma_1,\gamma_2,\gamma_3)$ of the form
$$\begin{cases}\gamma_1(t)=\dfrac{\cos\theta(\cos(kt)-1)-\sin\theta\sin(kt)}{k}\\ \gamma_2(t)=\dfrac{\sin\theta(\cos(kt)-1)+\cos\theta\sin(kt)}{k}\\ \gamma_3(t)=\dfrac{kt-\sin(kt)}{2k^2}\end{cases}. \tag{2.5.1}$$
Such geodesics are defined for $t\in[0,2\pi/|k|]$ and have length $2\pi/|k|$. When $k=0$, these curves degenerates to lines:
$$\begin{cases}\gamma_1(t)=-t\sin\theta\\ \gamma_2(t)=t\cos\theta\\ \gamma_3(t)=0\end{cases}.$$



(a) A geodesic with non-zero curvature in the Heisenberg geometry

(b) The isoperimetric sphere S_λ is C^∞ outside of the poles, and C^2 around them.

Figure 2.1: A ball and a geodesic in the Heisenberg geometry

We just found again the fact that lines through the origin in the xy-plane are geodesics.

Some consequences of the above characterization of the geodesics are the following facts. Denote by $C \subset \mathbb{R}^3$ the z-axis, which is the center of the Heisenberg group structure.

Fact 2.5.2. *For any $p \in \mathbb{R}^3 \setminus C$ there exists a unique geodesic arc connecting 0 and p.*

Fact 2.5.3. *For any $p \in C \setminus \{0\}$ there exists a one-parameter family of geodesic arcs connecting 0 and p.*

Fact 2.5.4. *The mapping $\Phi : \{(\theta, k, t) | \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in \mathbb{R}, t \in (0, \frac{2\pi}{|k|})\} \to \mathbb{R}^2 \setminus C$ given by*

$$\Phi(\theta, k, t) = \left(\frac{\cos\theta(\cos(kt) - 1) - \sin\theta\sin(kt)}{k}, \frac{\sin\theta(\cos(kt) - 1) + \cos\theta\sin(kt)}{k}, \frac{kt - \sin(kt)}{2k^2} \right) \tag{2.5.5}$$

is a homeomorphism.

Fact 2.5.6. *All the metric balls and metric spheres in the Heisenberg group are topological balls and spheres, respectively.*

We shall notice now that if $B(0, r)$ is the ball of center 0 and radius r, then

$$(x, y, z) \in B(0, 1) \Longleftrightarrow (rx, ry, r^2 z) \in B(0, r). \tag{2.5.7}$$

Indeed, if $(\gamma_1, \gamma_2, \gamma_3)$ is a geodesic arc of length 1 starting from the origin, then it is of the form (2.5.1) for some $k \in \mathbb{R}$ with $2\pi/|k| \geq 1$, and the time of the parametrization of (2.5.1) is $t \in [0, 1]$. Now the curve $(r\gamma_1, r\gamma_2, r^2\gamma_3)$ is

$$\left(\frac{\cos\theta(\cos(kt) - 1) - \sin\theta\sin(kt)}{k/r}, \frac{\sin\theta(\cos(kt) - 1) + \cos\theta\sin(kt)}{k/r}, \frac{kt - \sin(kt)}{2(k/r)^2} \right), \quad \text{for } t \in [0, 1],$$

which is a geodesic that is not parametrized by arc length, but by a multiple of it, namely r. Thus its length is r.

Notice that we did not use the homogeneous dilation $\mathbf{v} \mapsto r\mathbf{v}$; the third coordinate has been multiplied by r^2. Thus, such map $(x, y, z) \mapsto (rx, ry, r^2 z)$ multiplies the volume by a factor of r^4, and not r^3 as the usual Euclidean dilations do!

We can now deduce how is the growth of the balls in the Heisenberg geometry.

Proposition 2.5.8. *Let* Vol *be the 3D Lebesgue volume in $\mathbb{R}^3$, then*

$$\mathrm{Vol}(B(p, r)) = r^4 \,\mathrm{Vol}(B(p, 1)) = r^4 \,\mathrm{Vol}(B(0, 1)),$$

for all $r > 0$ and all $p \in \mathbb{R}^3$.

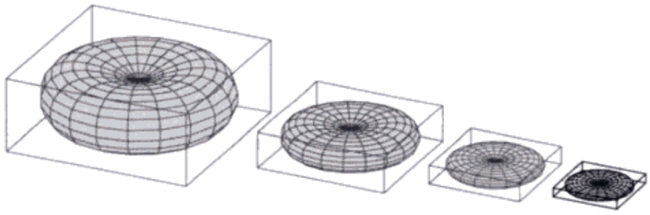
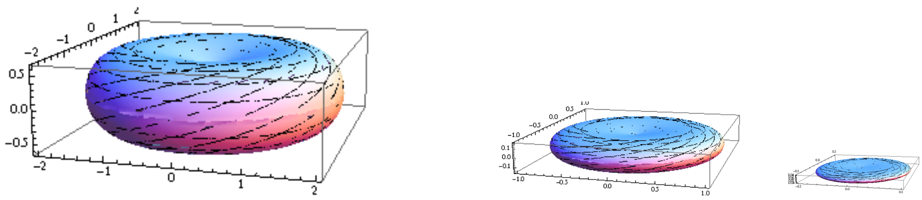


Figure 2.2: A picture by Bellaïche of balls of different sizes in the Heisenberg geometry.



(a) Radius 2 (b) Radius 1 (c) Radius 1/2

Figure 2.3: Balls of different sizes in the Heisenberg geometry

Proof. From (2.5.7) we know that $\mathrm{Vol}(B(0,r)) = r^4 \,\mathrm{Vol}(B(0,1))$. Now we can conclude the proof using both the fact that left translations (1.5.2) in the Heisenberg group are isometries together with the fact that they preserve the volume. This last fact can be checked noticing that the determinant of the differential of a left translations is 1, see (1.5.3). $\square$

Corollary 2.5.9. *The Heisenberg group endowed with the standard Carnot-Carathéodory distance has Hausdorff dimension equal to* 4.

Proof. It is enough to prove that there are positive constants k_1 and k_2 such that the minimal number N_ϵ of balls of radius ϵ needed to cover the unit ball satisfies

$$k_1 \epsilon^{-4} < N_\epsilon < k_2 \epsilon^{-4}.$$

For the lower bound, let $B_1, \dots, B_{N_\epsilon}$ be such balls. Then

$$\mathrm{Vol}(B(0,1)) \leq \sum_{j=1}^{N_\epsilon} \mathrm{Vol}(B_j) = N_\epsilon \epsilon^4 \,\mathrm{Vol}(B(0,1)).$$

For the upper bound, let $B_1, \dots, B_N$ be a maximal family of disjoint balls of radius $\epsilon/2$ in the unit ball. Then

$$\mathrm{Vol}(B(0,1)) \geq \sum_{j=1}^{N} \mathrm{Vol}(B_j) = N \left(\frac{\epsilon}{2}\right)^4 \mathrm{Vol}(B(0,1)).$$

Now, since the family $\{B_j\}_j$ is maximal, the ball with same centers as the B_j's and radius ϵ make up a cover of the unit ball. Thus

$$N_\epsilon \leq N \leq 16\epsilon^{-4}.$$

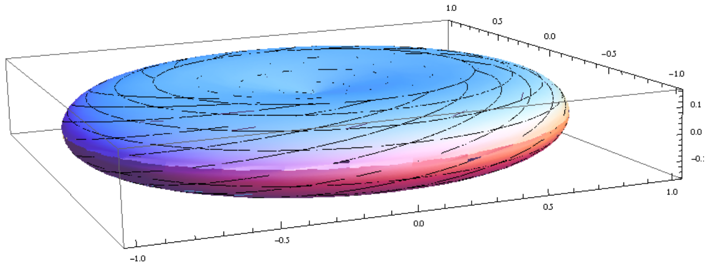


Figure 2.4: The unit ball in the Heisenberg geometry

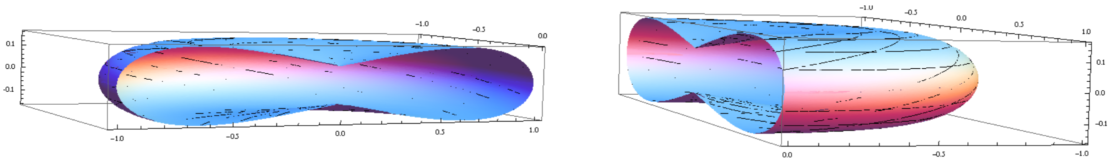


Figure 2.5: Sections of the unit ball in the Heisenberg geometry

□

More pictures

Here are some more pictures drawn with Mathematica 7.

One more picture should be added, showing that at the two "poles" the sphere is not smooth.

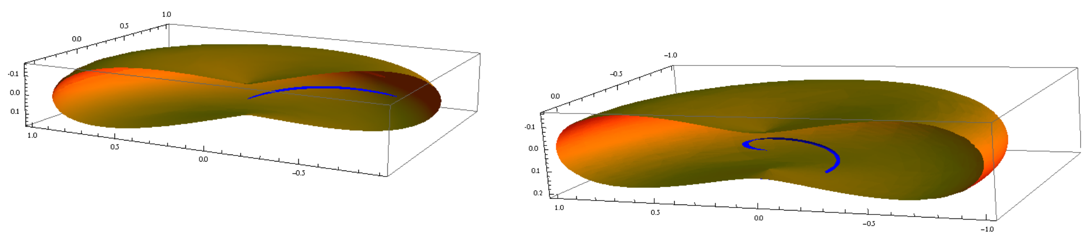


Figure 2.6: Balls and geodesics in the Heisenberg geometry

2.6 Ball-box Theorem and Hausdorff dimension

For the benefit of a clear exposition let us assume that a given distribution $\Delta \subset TM$ has the property that there exist vector fields $X_1, \ldots, X_n \in \Gamma(TM)$ such that, for all p in M, the first k

$$X_1(p), \ldots, X_k(p)$$

form a basis of $\Delta(p)$, they are a framing of M, i.e.,

$$X_1(p), \ldots, X_n(p)$$

form a basis of T_pM, and moreover, for each $j \in \{1, \ldots, n\}$, there exists $d_j \in \mathbb{N}$, called the *degree* of X_j, such that

$$X_j(p) \in \Delta^{[d_j]}(p) \setminus \Delta^{[d_j-1]}(p), \qquad \forall p \in M.$$

This last condition is a regularity assumption on Δ. We will call such Δ *equiregular distributions*. Not all distributions are equiregular, however, the following discussion can be generalized, see [Bel96].

The plan is to parametrize the manifold M using the flow of linear sums of such vector fields. Recall that, for $p \in M$ and $X \in \Gamma(TM)$, one denotes by $\exp_p(X)$ the value $\gamma(1)$ at time 1 of the integral curve γ of the vector field X starting at p, i.e., the solution of

$$\dot{\gamma} = X_\gamma \qquad \text{and} \qquad \gamma(0) = p.$$

Fixed $p \in M$, we define the exponential coordinates as

$$\Phi : \mathbb{R}^n \longrightarrow M$$

$$\Phi(t_1, \ldots, t_n) := \exp_p(t_1 X_1 + \ldots + t_n X_n).$$

Such map might be defined only on a neighborhood of $0 \in \mathbb{R}^n$.

The *box* with respect to $X_1, \ldots, X_n$ is defined as

$$\mathrm{Box}(r) := \{(t_1, \ldots, t_n) \in \mathbb{R}^n \ : \ |t_j| \leq r^{d_j}\}.$$

The following comparison theorem is due to many people (Mitchell, Gershkovich, Nagel-Stein-Wainger, cf. [Gro99]) and is called ball-box theorem since compare the boxes $\mathrm{Box}(r)$ in $\mathbb{R}^n$ with the balls $B(p, r)$ with respect to the d_{CC} distance.

Theorem 2.6.1 (Ball-Box Theorem)**.** *Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold with an equiregular distribution Δ. Let Φ be some exponential coordinate map from a point $p \in M$ constructed with respect to some equiregular basis $X_1, \ldots, X_n$. Then there are $C > 1$ and $\rho > 0$ such that*

$$\Phi\left(\mathrm{Box}(C^{-1}r)\right) \subseteq B(p,r) \subseteq \Phi\left(\mathrm{Box}(Cr)\right),$$

for all $r \in (0, \rho)$.

The Ball-Box Theorem will be proved later in the easier case of Carnot groups, cf Theorem 4.1.10.

Remark 2.6.2. The Ball-Box Theorem 2.6.1 gives a quantitative version of Chow's Theorems 2.4.1 and 2.4.2.

As far as we know, nothing is known regarding the following natural question, except for contact 3-manifolds.

Question 2.6.3 (Open!)**.** *Are all sufficiently small sub-Finsler balls and spheres homeomorphic to the usual Euclidean balls and spheres?*

Here is a first consequence of the Ball-Box Theorem 2.6.1.

Corollary 2.6.4 (Hölder equivalence of CC and Euclidean metrics)**.** *Locally, each sub-Finsler manifold is Hölder equivalent to a Rimannian manifold.*

Proof. Let $(M, \Delta, \|\cdot\|)$ be the sub-Finsler manifold. Let g be a Riemannian tensor whose norm is smaller than $\|\cdot\|$ and denote by d_{Riem} the induced Riemannian distance.

Consider the identity map $\mathrm{id} : M \to M$. Obviously the map

$$\mathrm{id} : (M, d_{CC}) \to (M, d_{Riem})$$

is 1-Lipschitz, and so Hölder.

Now, let $\alpha := \max_j d_j$ the maximum of the degree $\{d_j\}$ of the vector fields of some equiregular basis $\{X_j\}$. Using the facts that, for $\rho \in (0,1)$, one has that

$$\prod_{j=1}^{n}[-r^\alpha, r^\alpha] \subset \mathrm{Box}(r),$$

that the exponential maps has surjective differential at the origin, and the second inclusion of the Ball-Box Theorem 2.6.1, one shows that the map

$$\mathrm{id} : (M, d_{Riem}) \to (M, d_{CC})$$

is α-Hölder. $\square$

2.6.1 The problem of dimensions of hyper-surfaces in CC spaces

Define homogeneous dimension as

$$Q = \sum_{j=1}^{n} d_j, \tag{2.6.5}$$

which is in fact equal to (2.3.3).

Corollary 2.6.6. *If a sub-Finsler manifold $(M, \Delta, \|\cdot\|)$ has an equiregular distribution then the Hausdorff dimension of (M, d_{CC}) is the homogeneous dimension Q. Moreover, the Q-Hausdorff measure of (M, d_{CC}) is locally equivalent (up to multiplication by a function) of the Finsler volume form.*

Computing the Hausdorff dimension and Hausdorff measure of submanifolds in sub-Finsler manifolds with respect to the Carnot-Carathéodory distance is a rather natural question.

In 0.6 B of [Gro99], Gromov has given a general formula for the Hausdorff dimension of smooth submanifolds in equiregular Carnot-Carathéodory spaces and in [Mag08a] it is shown that this formula coincides with the degree of the submanifold, recently introduced in [MV08].

Theorem 2.6.7 ([Gro99, page104])**.** *Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold with an equiregular distribution Δ and Carnot-Carathéodory distance d_{CC}. Let $\Sigma \subset M$ a smooth sub-manifold. Then the Hausdorff dimension of (Σ, d_{CC}) is*

$$\dim_{Hau}(\Sigma, d_{CC}) = \max \left\{ \sum_{j=1}^{n} j \cdot \operatorname{rank}(T_pM \cap \Delta^{[j]}(p))/(T_pM \cap \Delta^{[j-1]}(p)) \ : \ p \in \Sigma \right\}.$$

Nevertheless, the question regarding Hausdorff measures of smooth submanifolds has not yet an answer. In [MV08] Magnani and Vittone found an integral formula for the spherical Hausdorff measure of submanifolds in Carnot groups under a suitable 'negligibility condition'. This negligibility condition has been recently obtained in all two step groups, [Mag08a] using standard covering arguments, and in the Engel group, using blow-up arguments [LDM08]. However it is still open in higher step groups and in general sub-Riemannian manifolds. We address the reader to the work of Magnani [MV08, Mag08b, Mag08a] for more information on this problem and its connections with the literature.

Problem 2.6.8. *Study Hölder equivalence and Hölder embeddings between sub-Riemannian manifolds.*

2.7 Exercises

Exercise 2.7.1. *Prove that any absolutely continuous curve in $\mathbb{R}^n$ can be re-parametrized to be a Lipschitz curve with respect to the Euclidean distance.*

Exercise 2.7.2. *Prove that Finsler-Carnot-Carathéodory distances, and in particular Riemannian and Finsler distances, are length distances.*

Exercise 2.7.3. *Let $(M, \Delta, \|\cdot\|)$ be a sub-Finsler manifold. We denote by $\mathrm{Length}_{d_{CC}}$ and $\mathrm{Length}_{\|\cdot\|}$ respectively the length with respect to the metric d_{CC} and the length with respect to the Finsler norm $\|\cdot\|$. Let γ be a horizontal curve. Show that*

$$\mathrm{Length}_{\|\cdot\|}(\gamma) = \mathrm{Length}_{d_{CC}}(\gamma).$$

Exercise 2.7.4. *Let γ be any absolutely continuous curve in a sub-Finsler manifold. Prove that*

$$\gamma \textit{ is horizontal } \Longleftrightarrow \mathrm{Length}_{d_{CC}}(\gamma) < +\infty.$$

Exercise 2.7.5. *Let $\Delta^{[j]}(p)$ the vector space defined in Section 2.3.1. Prove that*

$$\Delta^{[j]}(p) = \Delta^{[j-1]}(p) + \mathbb{R}\text{-}\mathrm{span}\left\{[X_1, [X_2[\ldots[X_{j-1}, X_j]]\ldots]](p) \ : \ X_1, \ldots X_j \in \Gamma(\Delta)\right\}.$$

Exercise 2.7.6. *Let $\Delta^{[j]}(p)$ the vector space defined in Section 2.3.1.*

1. *Show that $\Delta^{[j]}$ might not be a sub-bundle of TM.*

2. *Prove that, if $\Delta^{[j]}$ is a sub-bundle and so make sense to consider smooth sections $\Gamma(\Delta^{[j]})$ of the bundle $\Delta^{[j]}$, then*

$$\Delta^{[j+1]}(p) = \Delta^{[j]}(p) + \mathbb{R}\text{-}\mathrm{span}\left\{[X, Y](p) \ : \ X \in \Gamma(\Delta), Y \in \Gamma(\Delta^{[j]})\right\}.$$

Exercise 2.7.7. *Recall that $\Gamma(\Delta)$ denote the smooth sections of the bundle Δ. Define* $\mathrm{span}(\Delta) := \text{Lie-span}\{\Gamma(\Delta)\}$. *Show that the Hörmander's condition is equivalent to* $\mathrm{span}(\Delta) = TM$. (What is not immediately obvious is that elements of the form $[[X_1, X_2], [X_3, X_4]]$, with $X_1, X_2, X_3, X_4 \in \Gamma(\Delta)$, are contained in some $\Delta^{[j]}(p)$.)

Exercise 2.7.8. *Let Φ be the map defined in 2.5.5. Prova that the unit ball in the Heisenberg geometry is given by*

$$\begin{aligned} B(0,1) &= \{\Phi(\theta, k, t) | \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in \mathbb{R}, t \in (0,1)\} \\ &= \{\Phi(\theta, k, t) | \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in [-2\pi, 2\pi], t \in (0,1)\}, \end{aligned}$$

and the unit sphere is

$$S(0,1) = \{\Phi(\theta, k, 1) | \theta \in \mathbb{R}/2\pi\mathbb{Z}, k \in [-2\pi, 2\pi]\}.$$

Exercise 2.7.9. *Prove Corollary 2.6.6 from Theorem 2.6.1.*

Exercise 2.7.10. *Show, without using Theorem 2.6.7, that each smooth surface in the Heisenberg group has Haudorff dimension equal to* 3.

Exercise 2.7.11. *Give a proof of Corollary 2.6.6.*

Exercise 2.7.12. *Give a proof of Theorem 2.6.7.*

Chapter 3

Tangent cones and Carnot groups

Tangent spaces of a sub-Riemannian manifold are themselves sub-Riemannian manifolds. They can be defined as metric spaces, using Gromov's definition of tangent spaces to a metric space, and they turn out to be sub-Riemannian manifolds. Moreover, they come with an algebraic structure: nilpotent Lie groups with dilations. In the classical, Riemannian case, they are indeed vector spaces, that is, Abelian groups with dilations. Actually, the above is true only for regular points. At singular points, instead of nilpotent Lie groups one gets quotient spaces G/H of such groups G. Most of the exposition on tangent spaces is taken from [Bel96]. The prerequisites regarding Lie groups and Lie algebras are based on [War83] and [CG90].

3.1 Lie groups and their Lie algebras

Let $\mathbb{G}$ be a Lie group, i.e. a differentiable n-dimensional manifold with a smooth group operation.

Definition 3.1.1 (Lie group)**.** *A Lie group is a differentiable manifold*[1] *which is also endowed with a group structure such that the following map is* C^∞

$$\begin{array}{rcl} G \times G & \to & G \\ (x,y) & \mapsto & x^{-1}y. \end{array}$$

We shall denote by e the identity of the group, by $R_g(h) := hg$ the right translation, and by $L_g(h) := gh$ the left translation.

As in any manifold, the set of vector fields $\Gamma(TG)$ forms a Lie algebra. The general notion of Lie algebra is the following:

[1] Here we allow the case of non-connected manifolds, however we always consider second countability. Thus Lie groups can have at most countably many components.

Definition 3.1.2 (Lie algebra)**.** *A Lie algebra* $\mathfrak{g}$ *over* $\mathbb{R}$ *is a real vector space together with a bilinear operation*

$$[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$

(called the Lie bracket) such that, for all $x, y, z \in \mathfrak{g}$*,*

anti-commutativity $[x, y] = -[y, x]$

Jacobi identity $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0.$

The importance of the concept of Lie algebras is that there is a special finite dimensional Lie algebra intimately associated with each Lie group, and that properties of the Lie group are reflected in properties of its Lie algebra. We shall see, for example, that the connected, simply connected Lie groups are completely determined (up to isomorphism) by their Lie algebras.

The Lie algebra associated to a group is, as a vector space, the tangent T_eG at the identity. To see T_eG as a subset of $\Gamma(TG)$, we have to extend each vector to a vector field. Forced to make a choice, we follow the majority of the literature focusing on the *left* invariant vector fields. i.e. the vector fields $X \in \Gamma(T\mathbb{G})$ such that $(L_g)_*X = X$, so that $(dL_g)_xX = X_{L_g(x)}$ for all $x \in \mathbb{G}$. In differential terms, we have

$$X(f \circ L_g)(x) = Xf(L_g(x)) \qquad \forall x,\, g \in \mathbb{G}.$$

Thanks to (2.2.2) with $F = L_g$, the class of left-invariant vector fields is easily seen to be closed under the Lie bracket, and we shall denote by $\mathfrak{g} \subseteq \Gamma(T\mathbb{G})$ the *Lie algebra* of left-invariant vector fields.

Note that, after fixing a vector $v \in T_e\mathbb{G}$, we can construct a left-invariant vector field X defining $X_g := (L_g)_*v$ for any $g \in \mathbb{G}$. This construction is an isomorphism between the set $\mathfrak{g}$ of all left-invariant vector fields and $T_e\mathbb{G}$, and proves that $\mathfrak{g}$ is an n-dimensional subspace of $\Gamma(T\mathbb{G})$.

A map $F : G \to H$ between Lie groups is said a Lie group homomorphism is it is both C^∞ and a group homomorphism. A map $\phi : \mathfrak{g} \to \mathfrak{h}$ between Lie algebras is said a Lie algebra homomorphism is it is both linear and preserves brackets

$$\phi([X, Y]) = [\phi(X), \phi(Y)], \qquad \forall X, Y \in \mathfrak{g}.$$

The first connection between Lie groups and their Lie algebras is that each Lie group homomorphism indices a Lie algebra homomorphism: if $F : G \to H$ is a Lie group homomorphism, note that

$F(e) = e$, and one can easily show that the differential at the identity

$$dF_e : T_eG \to T_eH$$

preserves the bracket operation.

Vice versa, in the case when G is a Lie group that as a topological space is simply connected, then each Lie algebra homomorphism come from a Lie group homomorphism.

Theorem 3.1.3 ([War83, Theorem 3.27])**.** *Let G and H two Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively. Assume G simply connected. Let $\phi : \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra homomorphism. Then there exists a unique Lie group homomorphism $F : G \to H$ such that $dF = \phi$.*

Corollary 3.1.4. *If simply connected Lie groups G and H have isomorphic Lie algebras, then G and H are isomorphic.*

There is a theorem [Jac79, page 199] due to Ado that states that every Lie algebra has a faithful representation in $\mathfrak{gl}(n, \mathbb{R})$ for some n. As a consequence, if $\mathfrak{g}$ is a Lie algebra, then there exists a simply connected Lie group G with Lie algebra $\mathfrak{g}$. We then have the following correspondence.

Theorem 3.1.5. *There is a one-to-one correspondence between isomorphism classes of Lie algebras and isomorphism classes of simply connected Lie groups.*

3.1.1 Exponential map

Let M be any differentiable manifold. Let $X \in \Gamma(M)$ be a vector field. Fix a point $p \in M$ of the manifold. Then there is a unique curve $\gamma(t)$ satisfying $\gamma(0) = p$ with tangent $\dot{\gamma}(t) = X_{\gamma(t)}$. The corresponding exponential map is defined by $\exp_p(X) = \gamma(1)$. In general, the exponential map is only locally defined, that is, it only takes a small neighborhood of the zero section of TM, to a neighborhood of p in the manifold. This is because it relies on the theorem on existence and uniqueness of ordinary differential equation which is local in nature.

In the theory of Lie groups the exponential map is a map from the Lie algebra $\mathfrak{g}$ to the group G,

$$\exp\colon \mathfrak{g} \to G.$$

Elements of the Lie algebra $\mathfrak{g}$ are seen as left-invariant vector fields. Thus $\mathfrak{g} \subset \Gamma(TG)$ and so the previous definition make sense with $p = e$. Moreover, one can show that, for all $X \in \mathfrak{g}$, the ODE $\dot{\gamma}(t) = X_{\gamma(t)}$ has global solutions. Indeed, the curves $\gamma(t)$ are in this case homomorphisms from $\mathbb{R}$

to the group. Such homomorphisms from $\mathbb{R}$ to G are called one-parameter subgroups. Let $X \in \mathfrak{g}$ define the Lie algebra homomorphism

$$\varphi : T_0\mathbb{R} \to T_eG$$
$$t\partial_t \to tX.$$

Since $\mathbb{R}$ is simply connected, Theorem 3.1.3 asserts that there exists a one-parameter subgroup $\gamma : \mathbb{R} \to G$ with $d\gamma = \varphi$. This last condition just mean that $\dot\gamma(t) = X_{\gamma(t)}$. Indeed,

$$\begin{aligned}
\dot\gamma(t) &= \frac{d}{dh}\gamma(t+h)\Big|_{h=0} \\
&= \frac{d}{dh}\gamma(t)\gamma(h)\Big|_{h=0} \\
&= \frac{d}{dh}L_{\gamma(t)}(\gamma(h))\Big|_{h=0} \\
&= (L_{\gamma(t)})_*\dot\gamma(0) \\
&= (L_{\gamma(t)})_*(d\gamma)_0(\partial_t) \\
&= (L_{\gamma(t)})_*\varphi(\partial_t) \\
&= (L_{\gamma(t)})_*X \\
&= X_{\gamma(t)}.
\end{aligned}$$

We just proved the first part of point (iv) in the following theorem. In fact, the only non-trivial part of the theorem is point (iii) and the proof of it can be found in [War83, Theorem 3.31].

Theorem 3.1.6 ([War83, Theorem 3.31]). *Let $X \in \mathfrak{g}$ an element of the Lie algebra $\mathfrak{g}$ of a Lie group G.*

(i) $\exp((s+t)X) = \exp(sX)\cdot\exp(tX)$, *for $s, t \in \mathbb{R}$;*

(ii) $\exp(-X) = (\exp(X))^{-1}$;

(iii) $\exp : \mathfrak{g} \to G$ *is smooth and* $(d\exp)_0$ *is the identity map,*

$$(d\exp)_0 = \mathrm{id}_{\mathfrak{g}} : \mathfrak{g} \to \mathfrak{g},$$

so $\exp$ *gives a diffeomorphism of a neighborhood of* 0 *in* $\mathfrak{g}$ *onto a neighborhood of e in G;*

(iv) *The curve $\gamma(t) := \exp(tX)$ is the flow of X at time t starting from e, more generally, the curve $g\exp(tX) = L_g(\gamma(t))$ is the flow starting at g. As a particular consequence left-invariant vector fields are always complete.*

(v) *The flow of X at time t is the right translation $R_{\exp(tX)}$.*

Theorem 3.1.7 ([War83, Theorem 3.32])**.** *If $F : G \to H$ is a Lie group homomorphism, then*

$$F \circ \exp = \exp \circ dF.$$

The exponential map is in general different from the exponential map of Riemannian geometry. However, if G is compact, it has a Riemannian metric invariant under left and right translations, and the (Lie group) exponential map is the (Riemannian) exponential map of this Riemannian metric.

3.1.2 Nilpotent Lie groups and nilpotent Lie algebras

This material is taken from the book [CG90]. The exposition does not pretend to be a better one. We just extract for the book all those parts of importance for the following.

Let $\mathfrak{g}$ be a Lie algebra over $\mathbb{R}$. The lower (or descending) central series of $\mathfrak{g}$ is defined inductively by

$$\mathfrak{g}^{(1)} = \mathfrak{g};$$

$$\mathfrak{g}^{(i+1)} := [\mathfrak{g}; \mathfrak{g}^{(i)}] = \mathbb{R}\text{-span}\{[X, Y] : X \in \mathfrak{g}, Y \in \mathfrak{g}^{(i)}\}.$$

Definition 3.1.8. *(Nilpotent Lie algebra / Lie group). We say that $\mathfrak{g}$ is a nilpotent Lie algebra if there is an integer s such that $\mathfrak{g}^{(s+1)} = \{0\}$. Let s be the minimal integer such that $\mathfrak{g}^{(s+1)} = \{0\}$, then $\mathfrak{g}$ is said to be s-step nilpotent. A nilpotent Lie group is a Lie group whose Lie algebra is nilpotent.*

For connected Lie groups this is equivalent to saying that G itself is a nilpotent group.

A Lie algebra $\mathfrak{g}$ is s-step nilpotent if and only if all brackets of at least $s+1$ elements of $\mathfrak{g}$ are 0 but not all brackets of order s are.

Remark 3.1.9. A Lie algebra $\mathfrak{g}$ has always non-trivial center; in fact, if $\mathfrak{g}$ is s-step nilpotent, $\mathfrak{g}^{(s)}$ is central.

Recall that the *center* of a Lie algebra $\mathfrak{g}$ is

$$\mathrm{Center}(\mathfrak{g}) := \{X \in \mathfrak{g} \ : \ [X, Y] = 0 \text{ for all } Y \in \mathfrak{g}\},$$

and the *center* of a (Lie) group G is

$$\mathrm{Center}(G) := \{g \in G \ : \ gh = hg \text{ for all } h \in G\}.$$

The two centers are related since the center of a connected Lie group is a closed sub-group with Lie algebra the center of $\mathfrak{g}$, see [War83, page 116].

3.1.3 Examples

One common convention in describing nilpotent Lie algebras - and one that we shall often use - is the following. Suppose that $\mathfrak{g} = \mathbb{R}\text{-span}\{X_1, \dots, X_n\}$. To describe the Lie algebra structure of $\mathfrak{g}$, it suffices to give $[X_i, X_j]$ for all $i < j$. We can shorten this description considerably by giving only the non-zero brackets; all others are assumed to be zero.

Heisenberg algebras

The $(2n+1)$-dimensional Heisenberg algebra is the Lie algebra with basis $\{X_1, , \dots, X_n, Y_1, , \dots, Y_n, Z\}$, whose pairwise brackets are equal to zero expect for

$$[X_j, Y_j] = Z, \qquad \text{for } j = 1, \dots, n.$$

It is a two-step nilpotent Lie algebra. One way to realize it as a matrix algebra is to consider $(n+2) \times (n+2)$ upper triangular matrices of the form

$$\begin{pmatrix} 0 & x_1 & \dots & x_n & z \\ \cdot & 0 & \cdot & 0 & y_1 \\ . & & . & . & \vdots \\ \cdot & & & 0 & y_n \\ 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}.$$

The Lie group associated is called the n'th Heisenberg group and as matrix group it is

$$G = \left\{ \begin{pmatrix} 1 & x_1 & \dots & x_n & z \\ \cdot & 1 & \cdot & 0 & y_1 \\ . & & \ddots & . & \vdots \\ \cdot & & & 1 & y_n \\ 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix} : x_1, , \dots, x_n, y_1, , \dots, y_n, z \in \mathbb{R} \right\} \subset GL(n+2, \mathbb{R}).$$

Filiform algebras

The $(n+1)$-dimensional filiform algebra is the algebra spanned by $X, Y_1, Y_2, \dots, Y_n$, with only non-trivial relations

$$[X, Y_j] = Y_{j+1}, \qquad \text{for } j = 1, \dots, n-1.$$

It is an n-step nilpotent Lie algebra and can be realized as a matrix algebra considering the matrices of the form:

$$\begin{pmatrix} 0 & x & 0 & \cdot & 0 & y_n \\ & . & \ddots & & & \vdots \\ & & . & \ddots & & \vdots \\ & & & \cdot & x & y_2 \\ & & & & \cdot & y_1 \\ 0 & & & & & 0 \end{pmatrix}.$$

Strictly upper triangular matrix algebras

The algebra of strictly upper triangular $n \times n$ matrices is an $(n-1)$-step nilpotent Lie algebra of dimension $n(n-1)/2$, and its center is one-dimensional.

Free nilpotent algebras

The free nilpotent Lie algebra of step k and rank n (or on n generators) is defined to be the quotient algebra $\mathfrak{f}_n/\mathfrak{f}^{(k+1)}$, where $\mathfrak{f}_n$ is the free Lie algebra on n generators. It is not hard to see that it is finite-dimensional.

For example the Lie algebra of rank 2 and step 3 is given by the diagram

$$\begin{tikzcd}
X \arrow[dd] \arrow[dr] & & Y \arrow[dl] \arrow[dd] \\
& Z \arrow[dl] \arrow[dr] & \\
U & & V,
\end{tikzcd}$$

which has to be read as $[X, Y] = Z$, $[X, Z] = U$, and $[Z, Y] = V$.

3.1.4 The BCH formula

The Baker-Campbell-Hausdorff formula allows us to reconstruct any Lie group G locally, with its multiplication law, knowing only the structure of its Lie algebra $\mathfrak{g}$. The Baker-Campbell-Hausdorff formula links Lie groups to Lie algebras, by expressing the logarithm $\log(e^X e^Y)$ of the product of two Lie group elements as a Lie algebra element. The logarithm is by definition the inverse of the exponential, in general it is only locally defined in a neighborhood of the identity, thanks to Theorem 3.1.6(iii). However, for simply connected nilpotent Lie groups logarithm will be global by Theorem 3.1.13.

The general *Baker-Campbell-Hausdorff formula* (BCH formula, for short) is given by:

$$\log(\exp X \exp Y) = \sum_{n>0} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_i+s_i>0 \\ 1\le i\le n}} \frac{\left(\mathrm{ad}_X^{r_1} \circ \mathrm{ad}_Y^{s_1} \circ \mathrm{ad}_X^{r_2} \circ \mathrm{ad}_Y^{s_2} \ldots \circ \mathrm{ad}_X^{r_n} \circ \mathrm{ad}_Y^{s_n-1}\right)(Y)}{r_1!s_1!\cdots r_n!s_n!\ \sum_{i=1}^n (r_i+s_i)},$$

where $\mathrm{ad}_X Y = [X, Y]$, see (3.1.11). Thus

$$\left(\mathrm{ad}_X^{r_1} \circ \mathrm{ad}_Y^{s_1} \circ \mathrm{ad}_X^{r_2} \circ \mathrm{ad}_Y^{s_2} \ldots \circ \mathrm{ad}_X^{r_n} \circ \mathrm{ad}_Y^{s_n-1}\right)(Y)$$

$$= [\underbrace{X, [X, \ldots [X,}_{r_1} \underbrace{[Y, [Y, \ldots [Y,}_{s_1} \ldots \underbrace{[X, [X, \ldots [X,}_{r_n} \underbrace{[Y, [Y, \ldots Y}_{s_n}]] \ldots]].$$

The first terms of the series should[2] be

$$\begin{aligned}\log(\exp X \exp Y) &= X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] \\ &\quad -\frac{1}{24}[Y,[X,[X,Y]]] \\ &\quad -\frac{1}{720}([[[[X,Y],Y],Y],Y] + [[[[Y,X],X],X],X]) \\ &\quad +\frac{1}{360}([[[[X,Y],Y],Y],X] + [[[[Y,X],X],X],Y]) \\ &\quad +\frac{1}{120}([[[[Y,X],Y],X],Y] + [[[[X,Y],X],Y],X]) + \cdots\end{aligned}$$

3.1.5 Matrix groups

For matrix Lie groups $G \subseteq GL(n,\mathbb{R})$, the Lie algebra $\mathfrak{g} \subseteq \mathfrak{gl}(n,\mathbb{R})$ is simply the tangent space at the identity I with Lie bracket given by

$$[A,B] = AB - BA, \qquad \forall A, B \in \mathfrak{gl}(n,\mathbb{R}).$$

Moreover, the exponential map coincides with the exponential of matrices and is given by the ordinary series expansion:

$$\exp(A) = \sum_{j=0}^{\infty} \frac{1}{j!} A^j = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots, \qquad \forall A \in \mathfrak{gl}(n,\mathbb{R}). \tag{3.1.10}$$

(here I is the identity matrix). In this situation the Baker-Campbell-Hausdorff formula is obtained by formally solving for Z in $e^z = e^x e^y$,

$$\begin{aligned} Z &= \log(I + (e^X e^Y - I) \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (e^X e^Y - I)^n \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \left(\sum_{p_i+q_i>0, p_i, q_i>0} \frac{X^{p_i} Y^{q_i}}{p_i! q_i!} \right)^n \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sum_{p_i+q_i>0, p_i, q_i>0} \frac{X^{p_1} Y^{q_1} \cdots X^{p_n} Y^{q_n}}{p_1! q_2! \cdots p_n! q_n!}. \end{aligned}$$

One will get the BCH formula using that $\operatorname{ad}_A B = AB - BA$. Please, let me know if you find a clear and simple calculation of this ending.

3.1.6 Adjoint operators

Each Lie group acts on itself by conjugation: for $g \in G$, the map

$$C_g : h \mapsto ghg^{-1}$$

[2] I am not really sure of my calculations!

is an inner automorphism of G. Its differential at the unit element is called the *adjoin operator*:

$$\mathrm{Ad}_g = d(C_g)_e : \mathfrak{g} \to \mathfrak{g}.$$

The map Ad_g is a Lie algebra automorphism. For matrix groups we have the explicit formula:

$$\mathrm{Ad}_A(X) = AXA^{-1}, \qquad \text{for } A \in GL(n,\mathbb{R}) \text{ and } X \in \mathfrak{gl}(n,\mathbb{R}).$$

The action

$$\begin{array}{rcl} G \times \mathfrak{g} & \to & \mathfrak{g} \\ (g, X) & \mapsto & (\mathrm{Ad}_g)X \end{array}$$

is called the *adjoint action* of G. The map

$$\mathrm{Ad}_{(\cdot)} : G \to \mathrm{Aut}(\mathfrak{g})$$

is called the *adjoint representation* of G. Its differential ad $:= d(\mathrm{Ad})$is the adjoint map on $\mathfrak{g}$. One has that the following commutative diagram

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "\mathrm{ad}"] \arrow[d, "\exp"] & \mathrm{End}(\mathfrak{g}) \arrow[d, "\exp"] \\ G \arrow[r, "\mathrm{Ad}"] & \mathrm{Aut}(\mathfrak{g}) \end{tikzcd}$$

and the validity of the formula

$$\mathrm{ad}_X Y = [X, Y]. \tag{3.1.11}$$

Such maps satisfies the following formulae:

$$\exp((\mathrm{Ad}_g)Y) = C_g(\exp(Y)), \qquad \forall g \in G, Y \in \mathfrak{g},$$

$$C_{\exp(X)}(\exp(Y)) = \exp(\mathrm{Ad}_{\exp X}(Y)), \qquad \forall X, Y \in \mathfrak{g},$$

$$\mathrm{Ad}_{\exp X}(Y) = e^{\mathrm{ad}_X}(Y), \qquad \forall X, Y \in \mathfrak{g},$$

where

$$e^{\mathrm{ad}_X} := \sum_{j=0}^{\infty} \frac{1}{j!}(\mathrm{ad}_X)^j. \tag{3.1.12}$$

When G is a simply connected nilpotent Lie group the series 3.1.10 and 3.1.12 are finite, giving polynomial laws for the group multiplication and conjugation.

3.1.7 Simply connected nilpotent Lie groups

Recall from Corollary 3.1.4 that if two simply connected Lie groups have isomorphic Lie algebras, then they are isomorphic. We will see now how one can completely work on the Lie algebra using canonical coordinates.

Theorem 3.1.13 ([CG90, Theorem 1.2.1]). *Let G be a connected, simply connected nilpotent Lie group, with Lie algebra $\mathfrak{g}$. Then*

a *the exponential map* $\exp : \mathfrak{g} \to G$ *is an analytic diffeomorphism; and*

b *the Baker-Campbell-Hausdorff formula holds for all X and $Y \in \mathfrak{g}$.*

The following facts are consequences of the above theorem and its proof.

Fact 3.1.14. *Every Lie sub-group H of a connected, simply connected nilpotent Lie group G is closed and simply connected.*

Let N_n be the group whose Lie algebra are the strictly upper triangular matrices. Namely, N_n is the group of matrices that are upper triangular and have 1's in the diagonal.

Fact 3.1.15. *Every connected, simply connected nilpotent Lie group has a faithful embedding as a closed subgroup of N_n for some n.*

One important application of Theorem 3.1.13 involves coordinates on G. Since exp is a diffeomorphism of $\mathfrak{g}$ onto G, we can use it to transfer coordinates from $\mathfrak{g}$ to G. Some authors use exp to identify $\mathfrak{g}$ with G. Then the group multiplication can be calculated by the Baker-Campbell-Hausdorff formula.

Definition 3.1.16. *Let $\{X_1, \dots, X_n\}$ be any basis for a nilpotent Lie algebra $\mathfrak{g}$. The coordinates given by the map*

$$\Phi : \mathbb{R}^n \longrightarrow M$$

$$\Phi(t_1, \dots, t_n) := \exp(t_1 X_1 + \dots + t_n X_n)$$

are called exponential coordinates. Exponential coordinates are also known as canonical coordinates of the first kind.

Definition 3.1.17. *Let $\mathfrak{g}$ be a nilpotent Lie algebra. An ordered basis $\{X_1, \ldots, X_n\}$ for $\mathfrak{g}$ is called (strong) Malcev basis if, for each $k \in \{1, \ldots, n\}$, the space*

$$\mathbb{R}\text{-}\operatorname{span}\{X_1, \ldots, X_k\}$$

is an ideal[3] *of $\mathfrak{g}$*

Fact 3.1.18. *In the special class of Carnot groups, see next chapter, the existence of Malcev basis will be a triviality. However, any nilpotent algebra has Malcev basis, cf. Theorem 1.1.13 in [CG90] and the following notes.*

Lemma 3.1.19. *If $\{X_1, \ldots, X_n\}$ is a Malcev basis for a nilpotent Lie algebra $\mathfrak{g}$, then its ideals $\mathfrak{g}_k := \mathbb{R}\text{-}\operatorname{span}\{X_1, \ldots, X_k\}$ are such that*

$$[\mathfrak{g}, \mathfrak{g}_k] \subseteq \mathfrak{g}_{k-1}. \tag{3.1.20}$$

Proof. By definition of Malcev basis, we have $[\mathfrak{g}, \mathfrak{g}_k] \subseteq \mathfrak{g}_k$ and also $[\mathfrak{g}, \mathfrak{g}_{k-1}] \subseteq \mathfrak{g}_{k-1}$. If the conclusion of the lemma were not true, then there would be some $j \in \{1, \ldots, n\}$ and $a_1, \ldots, a_k$ with $a_k \neq 0$ such that

$$[X_j, X_k] = a_k X_k + \sum_{i=1}^{k-1} a_i X_i.$$

Now we iterate bracketing by X_j and we get, for some $a_1^{(l)}, \ldots, a_{k-1}^{(l)}$,

$$(\operatorname{ad}_{X_j}^l)(X_k) = a^l X_k + \sum_{i=1}^{k-1} a_i^{(l)} X_i,$$

which is never zero and so contradicts the nilpotency of $\mathfrak{g}$. $\square$

Definition 3.1.21. *Let $\{X_1, \ldots, X_n\}$ be a (strong) Malcev basis for a nilpotent Lie algebra. Define the map*

$$\Psi : \mathbb{R}^n \to G$$

$$\Psi(s) := \exp(s_1 X_1) \cdots \exp(s_n X_n).$$

The coordinate system defined is called strong Malcev coordinates *or also* canonical coordinates of the second kind.

[3] A subspace $I \subseteq \mathfrak{g}$ is called an ideal of $\mathfrak{g}$ if $[\mathfrak{g}, I] \subseteq I$. By anticommutativity, there is no need of distinction between left and right ideals.

If $\{X_1, \dots, X_n\}$ is a Malcev basis for a nilpotent Lie algebra, we can consider both canonical coordinates; we have that the Malcev coordinates are related to the exponential coordinates by a polynomial diffeomorphism whose Jacobian determinant is constantly equal to 1.

Proposition 3.1.22 ([CG90, Proposition 1.2.7])**.** *Let $\{X_1, \dots, X_n\}$ be a Malcev basis for a nilpotent Lie algebra $\mathfrak{g}$. Let $\Psi : \mathbb{R}^n \to G$ the Malcev coordinate system and $\Phi : \mathbb{R}^n \longrightarrow M$ the exponential coordinate system associated to the basis. Then*

(i) *$\Psi(s) = \Phi(P(s))$ where $P : \mathbb{R}^n \to \mathbb{R}^n$ is a polynomial diffeomorphism with polynomial inverse.*

(ii) *writing $P = (P_1, \dots, P, n)$, then $P_j(s) = s_j + \hat{P}(s_{j+1}, \dots, s_n)$.*

In other words, we have the relation:

$$\exp(s_1 X_1) \cdots \exp(s_n X_n) = \exp(P_1(s) X_1 + \ldots + P_n(s) X_n).$$

Proposition 3.1.23 ([CG90, Proposition 1.2.9])**.** *Assume that G is equipped with either exponential or Malcev coordinates with respect to some basis. For any $g \in G$, left translation L_g and right translation R_g are maps whose Jacobian determinants are identically equal to* 1.

Proof. We prove the statement for exponential coordinates and left translations. The case of right translations is similar. For Malcev coordinates it will be true because of they differs from exponential coordinates by a polynomial diffeomorphism whose Jacobian determinant is constantly equal to 1, Proposition 3.1.22.

The proof is based on the BCH formula and (3.1.20). Indeed, we can assume that the basis $\{X_1, \dots, X_n\}$ is a Malcev basis, since linear changes of basis preserve Jacobians. So, let Φ the exponential coordinate system, and L_g the left translation by g. We need to calculate the Jacobian of $\Phi^{-1} \circ L_g \circ \Phi$. Thus we consider the diagram

$$\begin{array}{ccc}
\mathbb{R}^n & \xrightarrow{\quad\quad\Phi\quad\quad} & G \\
 & & \\
(t_1, \dots, t_n) & \longmapsto & \exp(\sum_j t_j X_j) \\
 & & \big\downarrow {\scriptstyle L_g} \\
(s_1, \dots, s_n) & \longmapsto & g \exp(\sum_j t_j X_j) \\
 & & \\
\mathbb{R}^n & \xrightarrow{\quad\quad\Phi\quad\quad} & G,
\end{array}$$

and we solve the dependence of the s_i's from the t_j's. Since the Malcev coordinates are surjective we can find $u_1, \dots, u_n$ and write

$$g = \exp(u_1 X_1) \dots \exp(u_n X_n).$$

It is enough to consider the case $g = \exp(u_k X_k)$ and then conclude considering compositions. Thus we need to consider the system

$$\exp(\sum_j s_i X_i) = \exp(u_k X_k) \exp(\sum_j t_j X_j).$$

By the BCH formula,

$$\sum_j s_i X_i = u_k X_k + \sum_j t_j X_j + \frac{1}{2}[u_k X_k, \sum_j t_j X_j] + \dots.$$

Since we have chosen a Malcev basis we have the property (3.1.20). Thus a bracket as $[X_k, X_j]$ is only a combination of $\{X_1, \dots, X_{j-1}\}$. In other words, the function s_j is of the form t_j plus a polynomial that does not depend on the variables $t_1, \dots, t_j$. In coordinates, the map $\Phi^{-1} \circ L_g \circ \Phi$ is of the form

$$\Phi^{-1} \circ L_g \circ \Phi = \begin{pmatrix} t_1 & * & \dots & \dots & \dots & \dots & * \\ 0 & \ddots & \ddots & * & \dots & * & \vdots \\ \cdot & \cdot & t_{k-1} & * & \ddots & \vdots & \vdots \\ \cdot & \cdot & 0 & t_k + u_k & * & * & \vdots \\ \cdot & \cdot & \cdot & 0 & t_{k-1} & \ddots & \vdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \ddots & * \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & t_n \end{pmatrix}.$$

Thus the differential is of the form

$$d(\Phi^{-1} \circ L_g \circ \Phi) = \begin{pmatrix} 1 & * & \dots & \dots & \dots & \dots & * \\ 0 & \ddots & \ddots & * & \dots & * & \vdots \\ \cdot & \cdot & 1 & * & \ddots & \vdots & \vdots \\ \cdot & \cdot & 0 & 1 & * & * & \vdots \\ \cdot & \cdot & \cdot & 0 & 1 & \ddots & \vdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \ddots & * \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}.$$

Thus the Jacobian of a left translation in exponential coordinates with respect to a Malcev basis is 1 at every point. $\square$

All the map of the form $\exp(P_1(s)X_1 + \dots + P_n(s)X_n)$, where $P : \mathbb{R}^n \to \mathbb{R}^n$ is a polynomial diffeomorphism with polynomial inverse, are called *polynomial coordinate map*. Examples are, obviously, exponential and, by Proposition 3.1.22, Malcev coordinate maps.

The key observation is that the Jacobian of any polynomial diffeomorphism with polynomial inverse is a polynomial that is invertible inside the polynomial ring, so it is a constant. Thus, changing of coordinates by a polynomial diffeomorphism with polynomial inverse preserves Lebesgue measure preserving maps.

Any Lie group, as any locally compact group, has a natural class of measures: the *Haar measures.* A measure μ is called a left-Haar measure if it is left-invariant, i.e., if, for any left translation L_g,

$$((L_g)_{\#}\mu)(B) := \mu\left(L_g^{-1}(B)\right) = \mu(B), \qquad \text{for all Borel set } B.$$

Left-Haar measures, as right-Haar measures, are unique up to a multiplication by a constant.

A consequence of the previous proposition and the last observation above is the following theorem.

Theorem 3.1.24 ([CG90, Theorem 1.2.10])**.** *Let G be an n-dimensional connected, simply connected, and nilpotent Lie group. Any polynomial coordinate map pushes forward the Lebesgue measure on $\mathbb{R}^n$ to a Haar measure on G.*

It is not always true that left-Haar measures are also right-Haar measures, groups with such property are called *unimodular.* However in any nilpotent Lie group Haar measure are both left and right-invariant. Theorem 3.1.24 shows such uniqueness for simply connected nilpotent Lie groups and it is suffices for our cases of interest.

3.1.8 Homogeneous manifolds

This part will probably will omitted in class.

Theorem 3.1.25 ([War83, Theorem 3.58])**.** *Let [...]*

Theorem 3.1.26 ([CG90, Theorem 1.2.12])**.** *Let [...]*

Theorem 3.1.27 ([CG90, Theorem 1.2.13])**.** *Let [...]*

read page 23 [CG90] remark 1 and 3.

3.2 Carnot groups

A Carnot group $\mathbb{G}$ of step $s \geq 1$ is a connected, simply connected Lie group whose Lie algebra $\mathfrak{g}$ admits a step s stratification: this means that we can write

$$\mathfrak{g} = V_1 \oplus \cdots \oplus V_s,$$

with $[V_j, V_1] = V_{j+1}$, for $1 \leq j \leq s-1$, and $V_s \neq \{0\}$.

Question 3.2.1. *Can a Carnot group have two non-isomorphic stratifications of its Lie algebra? no!*

Not all simply connected Lie groups that are nilpotent are in fact Lie groups: there are 6-dimensional nilpotent Lie algebras that are not stratified, cf. [Goo76].

We shall keep the notation $n = \sum_i \dim V_i$ for the topological dimension of $\mathbb{G}$. The value Q defined by

$$Q := \sum_{i=1}^{s} i \dim V_i \tag{3.2.2}$$

will be the so-called *homogeneous dimension* of $\mathbb{G}$ as a sub-Riemannian manifold.

Indeed, each Carnot group is intrinsically equipped with a sub-Riemannian structure, which is unique up to biLipschitz equivalence and has an additional structure of dilations, which we will explain in the next subsection. Fixed the stratification, let Δ be the left-invariant sub-bundle of the tangent bundle TG with $\Delta_e = V_1$. Let $\|\cdot\|$ be any left-invariant Finsler norm on G, equivalently one has to fix a Banach norm on $\mathfrak{g} = T_e G$ and extend it by left-translations. The triple $(M, \Delta, \|\cdot\|)$ is now a sub-Finsler manifold which satisfies the Hörmander's condition, since one has that

$$\Delta^{[j]}(e) = V_1 \oplus \cdots \oplus V_j.$$

Thus one can consider the Finsler-Carnot-Carathéodory distance d_{CC} associated to the sub-Finsler structure.

One should immediately observe that another choice of the norm would not change the biLipschitz equivalence class of the sub-Finsler manifold. Namely, if $\|\cdot\|_2$ is another left-invariant Finsler norm on G, then

$$\mathrm{id} : (\mathbb{G}, d_{CC,\|\cdot\|}) \to (\mathbb{G}, d_{CC,\|\cdot\|_2})$$

is globally biLipschitz. So as a consequence of our interest to metric spaces up to biLipschitz equivalence, we may assume that the norm $\|\cdot\|$ is coming from a scalar product $\langle\cdot|\cdot\rangle$.

In the definition of the Carnot-Carathéodory distance only the value of the scalar product on V_1, and not on all $\mathfrak{g}$, is important. Defining a scalar product on V_1 is equivalent to specifying an orthonormal basis of it. So, denoting by m the dimension of V_1, we fix an inner product in V_1 by fixing an orthonormal basis $X_1, \ldots, X_m$ of V_1. This basis of V_1 induces the Carnot-Carathéodory left-invariant distance d in $\mathbb{G}$, which we recall can be defined as follows:

$$d(x,y) := \inf \left\{ \int_0^1 \sqrt{\sum_{i=1}^m |a_i(t)|^2}\, dt : \ \gamma(0)=x,\ \ \gamma(1)=y \right\},$$

where the infimum is among all absolute continuous curves $\gamma : [0,1] \to \mathbb{G}$ such that $\dot\gamma(t) = \sum_1^m a_i(t)(X_i)_{\gamma(t)}$ for a.e. $t \in [0,1]$ (the so-called horizontal curves).

3.2.1 The dilation structure

The construction of the dilation structure deeply uses the stratification of the algebra: $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$. We denote by $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ the family of inhomogeneous dilations defined by

$$\delta_\lambda \left(\sum_{i=1}^s v_i \right) := \sum_{i=1}^s \lambda^i v_i \qquad \lambda \geq 0$$

where $X = \sum\limits_{i=1}^s v_i$ with $v_i \in V_i$, $1 \leq i \leq s$. The dilations δ_λ belong to $GL(\mathfrak{g})$ and are uniquely determined by the homogeneity conditions

$$\delta_\lambda X = \lambda^k X \quad \forall X \in V_k, \qquad 1 \leq k \leq s.$$

By Theorem 3.1.13, in Carnot groups the map $\exp : \mathfrak{g} \to \mathbb{G}$ is a diffeomorphism, so any element $g \in \mathbb{G}$ can represented as $\exp(X)$ for some unique $X \in \mathfrak{g}$, and therefore uniquely written in the form

$$\exp\left(\sum_{i=1}^s v_i \right), \qquad v_i \in V_i,\ 1 \leq i \leq s. \tag{3.2.3}$$

This representation allows to define a family indexed by $\lambda \geq 0$ of intrinsic dilations $\delta_\lambda : \mathbb{G} \to \mathbb{G}$, by

$$\delta_\lambda \left(\exp\left(\sum_{i=1}^s v_i \right) \right) := \exp\left(\sum_{i=1}^s \lambda^i v_i \right) \qquad \text{(i.e. } \exp \circ \delta_\lambda = \delta_\lambda \circ \exp.)$$

We have kept the same notation δ_λ for both dilations (in $\mathfrak{g}$ and in $\mathbb{G}$) because no ambiguity will arise. Obviously, $\delta_\lambda \circ \delta_\eta = \delta_{\lambda\eta}$, and the Baker-Campbell-Hausdorff formula gives

$$\delta_\lambda(xy) = \delta_\lambda(x)\delta_\lambda(y) \qquad \forall x,\, y \in \mathbb{G}. \tag{3.2.4}$$

Another useful relation between dilations in $\mathbb{G}$ and dilations in $\mathfrak{g}$ is $\delta_\lambda X = (\delta_\lambda)_* X$, namely

$$X(u \circ \delta_\lambda)(g) = (\delta_\lambda X)u(\delta_\lambda g) \qquad \forall g \in \mathbb{G},\ \lambda \geq 0. \tag{3.2.5}$$

We have indeed

$$\begin{aligned} X(u \circ \delta_\lambda)(g) &= \frac{d}{dt} u \circ \delta_\lambda(g \exp(tX))\Big|_{t=0} = \frac{d}{dt} u(\delta_\lambda g \delta_\lambda \exp(tX))\Big|_{t=0} \\ &= \frac{d}{dt} u(\delta_\lambda g \exp(t\delta_\lambda X))\Big|_{t=0} = (\delta_\lambda X)u(\delta_\lambda g). \end{aligned}$$

Another viewpoint is that $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ is a Lie algebra isomorphism since, from the grading, it is obvious that

$$\delta_\lambda([X,Y]) = [\delta_\lambda X, \delta_\lambda Y].$$

Then the map $\delta_\lambda : \mathbb{G} \to \mathbb{G}$ is the unique group homomorphism with δ_λ as differential, whose existence is given by Theorem 3.1.3 since $\mathbb{G}$ is simply connected.

Moreover, the Carnot-Carathéodory distance is well-behaved under these dilations, namely

$$d(\delta_\lambda x, \delta_\lambda y) = \lambda d(x,y) \qquad \forall x,\, y \in \mathbb{G}. \tag{3.2.6}$$

Indeed, if γ in a horizontal curve from x to y, then $\delta_\lambda \circ \gamma$ is a curve going from $\delta_\lambda x$ to $\delta_\lambda y$ whose tangent vectors are, for almost all t,

$$(\delta_\lambda)_* \dot\gamma(t) = \delta_\lambda(\dot\gamma(t)) = \lambda \dot\gamma(t), \tag{3.2.7}$$

which are horizontal since $\dot\gamma(t)$ is horizontal. Moreover, from (3.2.7), the length of $\delta_\lambda \circ \gamma$ is λ times the length of γ. Thus 3.2.6 has been shown.

3.3 Nilpotization

We explain now what is the Carnot group which appear as tangent to a given equi-regular distribution. Let Δ be a bracket-generating and equi-regular distribution in a manifold M, i.e.,

$$\Delta = \Delta^{[1]} \subset \Delta^{[2]} \subset \ldots \subset \Delta^{[s]} = TM$$

is a flag of sub-bundles of TM, where $\Delta^{[j+1]} = \Delta^{[j]} + [\Delta, \Delta^{[j]}]$. Note that in the last sum is not necessarily a direct sum. The simple but crucial fact is that

$$[\Delta^{[k]}, \Delta^{[l]}] \subseteq \Delta^{[k+l]}. \tag{3.3.1}$$

Equation (3.3.1) is obvious for $k=1$ and can be proved by induction using Jacobi identity:

$$\begin{aligned}
[\Delta^{[k+1]}, \Delta^{[l]}] &= \Big[\Delta^{[k]} + [\Delta, \Delta^{[k]}], \Delta^{[l]}\Big] \\
&= [\Delta^{[k]}, \Delta^{[l]}] + \Big[[\Delta, \Delta^{[k]}], \Delta^{[l]}\Big] \\
&\subseteq \Delta^{[k+l]} + \Big[[\Delta^{[k]}, \Delta^{[l]}], \Delta\Big] + \Big[[\Delta^{[l]}, \Delta], \Delta^{[k]}\Big] \\
&\subseteq \Delta^{[k+l]} + [\Delta^{[k+l]}, \Delta] + [\Delta^{[l+1]}, \Delta^{[k]}] \\
&\subseteq \Delta^{[k+l]} + \Delta^{[k+l+1]} + \Delta^{[k+l+1]} \\
&\subseteq \Delta^{[k+l+1]}
\end{aligned}$$

Define $H_1 := \Delta$ and $H_j := \Delta^{[j]}/\Delta^{[j-1]}$, for $j = 2, \ldots, n$. Still H_j is a bundle over M, but not a sub-bundle of the tangent bundle TM. We obviously have the following isomorphism

$$TM \simeq H_1 \oplus H_2 \oplus \ldots \oplus H_s.$$

In this notes we also assume that the equi-regular distributions have the further property of having a global framing $X_1, \ldots, X_n$ of M such that, for some $m_1, \ldots, m_s$,

$$\Delta^{[j]}(p) = \mathbb{R}\text{-}\operatorname{span}\{X_1(p), \ldots, X_{m_j}(p)\}, \qquad \forall p \in M.$$

Fact 3.3.2. *For each point $p \in M$, the vector space T_pM inherits the structure of a Carnot group, with respect the stratification $H_j(p)$. Such Carnot group is sometimes called the* nilpotization *of T_pM with respect to Δ.*

The following proof is incomplete - a new proof will be given in the future - for now see [Bul02]. Let $V_j := H_j(p)$. Obviously T_pM and $V_1 \oplus \cdots \oplus V_s$ are isomorphic vector spaces. We need to define a Lie algebra product and then show that $[V_j, V_1] = V_{j+1}$. Take $x, y \in T_pM$, with $x \in V_j$ and $y \in V_l$. Since $V_j = H_j(p) = \Delta^{[j]}(p)/\Delta^{[j-1]}(p)$, we have that there exist $X \in \Delta^{[j]}$ and $Y \in \Delta^{[l]}$, such that

$$x = X(p) + \Delta^{[j-1]}(p) \qquad \text{and} \qquad y = Y(p) + \Delta^{[l-1]}(p).$$

We define, naturally,

$$[x, y] := [X, Y](p) + \Delta^{[j+l-1]}(p).$$

The definition is well posed because of (3.3.1): if $u \in \Delta^{[j-1]}$, then $[X + u, Y] = [X, Y] + [u, Y]$, with $[u, Y] \in [\Delta^{[j-1]}, \Delta^{[l]}] \subseteq \Delta^{[j+l-1]}$. Thus $[X + u, Y](p)$ and $[X, Y](p)$ are equal mod $\Delta^{[j+l-1]}(p)$. NEED TO SHOW INDEPENDENCE FROM THE REPRESENTATIVE X.

Again, if $y \in V_1$, from (3.3.1) we immediately have that $[x, y] \in \Delta^{[j+1]}(p)/\Delta^{[j]}(p) = V_{j+1}$. Thus $[V_j, V_1] \subseteq V_{j+1}$. To show the reverse inclusion, let $z \in V_{j+1}$. Consider a representative $Z \in \Delta^{[j+1]}$ such that $z = Z(p) + \Delta^{[j]}(p)$. By definition $\Delta^{[j+1]} = \Delta^{[j]} + [\Delta^{[j]}, \Delta]$, so there are $W \in \Delta^{[j]}$, $X_l \in \Delta^{[j]}$, and $Y_l \in \Delta$ such that $Z = W + \sum_l [X_l, Y_l]$. Take $x_l = X_l(p) \pmod{\Delta^{[j-1]}}$ and $y_l = Y_l(p)$. We have then

$$\begin{aligned}
\sum_l [x_l, y_l] &= \sum_l [X_l, Y_l](p) \qquad \pmod{\Delta^{[j]}(p)} \\
&= (Z - W)(p) \qquad \pmod{\Delta^{[j]}(p)} \\
&= Z(p) \qquad \pmod{\Delta^{[j]}(p)}.
\end{aligned}$$

Therefore we have shown that $[V_j, V_1] = V_{j+1}$. $\square$

3.4 Mitchell's Theorem on tangent cones

Given a metric space (X, d), one defines the dilated metric space $(X, \lambda d)$ dilated by a factor of $\lambda \in \mathbb{R}$ as the same set X endowed with the dilated distance $(\lambda d)(p, q) := \lambda d(p, q)$. Gromov has defined the notion of tangent space to a metric space as limit of such objects.

We say that a metric space (Z, ρ) is *a* tangent of (X, d) at the point $p \in X$ if there exists $\bar{p} \in Z$ and a sequence $\lambda_j \to \infty$ such that

$$\lim_j (X, p, \lambda_j d) = (Z, \bar{p}, \rho).$$

It signifies [4] that for each $r > 0$, there is a sequence of $\epsilon_j \to 0$ such that the ball of radius $r + \epsilon_j$ in $(X, \lambda_j d)$ about the base point p converges to the ball of radius r about $\bar{p}$. Namely, the infimum of the Gromov-Hausdorff distance between these compact abstract metric spaces approach 0 as $\lambda_j \to \infty$.

The Gromov-Hausdorff distance $\mathrm{GH}(B_1, B_2)$ between two compact metric spaces B_1 and B_2 is infimum $\inf_{\psi_1,\psi_2} \mathrm{H}(\psi_1 B_1, \psi_2 B_2)$ over all isometric embeddings ψ_1, ψ_2 of B_1 and B_2 into the same metric space C of the Hausdorff distance $\mathrm{H}(\psi_1 B_1, \psi_2 B_2)$ of the images as subset of C.

A distribution is said to be generic if, for each j, $\dim \Delta^{[j]}(p)$ is independent of the point p in M.

Theorem 3.4.1 (Mitchell)**.** *For a generic distribution Δ on M, the tangent cone of a sub-Riemannian manifold (M, d_{CC}) at $p \in M$ is isometric to (G, d_∞) where G is a Carnot group with a left-invariant Carnot-Carathéodory metric. In fact, the group G is the nilpotization of T_pM with respect to Δ.*

Remark 3.4.2. The simple fact that we would like the reader to observe is that the tangent cone of a Carnot group $\mathbb{G}$ is $\mathbb{G}$ itself. Indeed, dilations δ_λ provide isometries between (G, d_{CC}) and $(G, \lambda d_{CC})$.

Remark 3.4.3. Differently from the Riemannian case, it is NOT true that a sub-Riemannian manifold is locally biLipschitz equivalent to its tangent cone. It is however true for contact manifolds because of Darboux Theorem.

3.5 Pansu's Rademacher Theorem

We would like to observe that the classical Rademacher Theorem states not only the existence almost everywhere of a tangent map (called the differential), but also its realizability as a linear map, in

[4] In the case when the metric space (X, d) is geodesic, the limit should be easier to understand. Look at [BBI01, page 272].

other word, as a group homomorphism which is compatible with the respective groups of dilations. Stated in this terms, the theorem holds for general sub-Riemannian manifolds too.

Theorem 3.5.1 (Pansu's Rademacher Theorem [Pan89, MM95])**.** *At almost all points, the tangent map of a Lipschitz map between sub-Riemannian manifolds exists, is unique, and is a group homomorphism of the tangent cones equivariant with respect to their dilations.*

Let us clarify what is the tangent map. Each map $f : (X, d) \to (X', d')$ induces a map $f_\lambda : (X, \lambda d) \to (X', \lambda d')$, for each $\lambda > 0$, which set-wise is the same map $f(x) = f_\lambda(x)$. Fix a point $x \in X$ and assume that (Z, ρ) and (Z', ρ') are tangent cone respectively to (X, d) at x and to (X', d') at $f(x)$. One says that $\hat{f} : (Z, \rho) \to (Z', \rho')$ is a *tangent map* of f at x if, for some sequence $\lambda_j \to \infty$, f_{λ_j} converges uniformly on compact sets to $\hat{f}$.

With such a definition above the tangent map could not be unique and could not be linear. For example the absolute value map has itself as tangent at zero. For Lipschitz maps between sub-Riemannian manifolds, Pansu-Rademacher Theorem 3.5.1 states not only existence at almost every point of such a blow up map, but also that, outside of a negiglible set, such limit is a Lie group homomorphism between Carnot groups that commutes with Carnot group dilations.

The tangent map could be also called blow up differential or, in the Carnot group setting, Pansu differential. If $f : \mathbb{G}_1 \to \mathbb{G}_2$ is a map between two Carnot groups with dilations $\delta^{(1)}_\lambda$ and $\delta^{(2)}_\lambda$ respectively, then its Pansu differential at a point x is the limit

$$\lim_{\lambda \to \infty} \delta^{(2)}_\lambda \circ L^{-1}_{f(x)} \circ f \circ L_x \circ \delta^{(1)}_{1/\lambda}.$$

Let us warn the reader about a possible confusion. Each sub-Riemannian manifold is in particular a differentiable manifold. However, the notion of the differential of a smooth map does not coincide with the tangent map which is defined in geometric terms.

3.5.1 Applications to non-embeddability

It was observed by Semmes, [Sem96, Theorem 7.1], that Pansu's differentiation Theorem 3.5.1 implies that a Lipschitz embedding of the Heisenberg group with its CC distance into an Euclidean space, cannot be bi-Lipschitz.

Theorem 3.5.2. *There is no bi-Lipschitz embedding from an open set in a Heisenberg group to an Euclidean space $\mathbb{R}^n$.*

Proof. Suppose that such an embedding f exists. The Pansu Rademacher Theorem 3.5.1 would imply that there exists at least one point at which f is differentiable and whose tangent map is a group homomorphism. The blowing-up procedure used to define the tangent map scales in the natural way, i.e., if f is L-bi-Lipschitz, then each rescaled f_λ is L-bi-Lipschitz and so the tangent map is bilipschitz too. In particular, the tangent map is injective. We now get a contradiction, because we considered a tangent map which is a group homomorphism between tangents spaces which are the 3-dimensional Heisenberg group and the Abelian $\mathbb{R}^n$. However, any homomorphism from the Heisenberg group into $\mathbb{R}^n$ must have a kernel which is at least 1-dimensional (all commutators in the Heisenberg group must be mapped to 0 by the homomorphism) and hence cannot be injective. □

Corollary 3.5.3. *Let M_1 an M_2 be sub-Riemannian manifolds with tangents the Carnot groups $\mathbb{G}_1$, respectively $\mathbb{G}_2$. If no subgroup of $\mathbb{G}_2$ is isomorphic to $\mathbb{G}_1$ then there is no bi-Lipschitz embedding of M_1 in M_2.*

Corollary 3.5.4. *The Heisenberg group, or any other non-commutative Carnot group, is purely unrectifiable.*

A consequence of the proof of Theorem 3.5.2 is that each Lipschitz map from the Heisenberg group to an Euclidean space has to compress points in the direction of the center of the group.

Proposition 3.5.5 (Center collapse)**.** *If $U \subset H$ is an open subset, and $f : U \to \mathbb{R}^n$ is a Lipschitz map, then for almost every point $x \in H$, the map collapses in the direction of the center of H, i.e.*

$$\lim_{g \to e} \frac{\|f(gx) - f(x)\|}{d(gx, x)} = 0\,, \qquad g \in \mathrm{Center}(H)\,. \tag{3.5.6}$$

This last theorem has been generalized by J. Cheeger and B. Kleiner to maps with values in the Banach space L^1. Such a result gave a proof of the following theorem which has been conjectured by J. Lee and A. Naor.

Theorem 3.5.7 (Lee-Naor-Cheeger-Kleiner)**.** *The Heisenberg group equipped with its CC metric does not admit a bi-Lipschitz embedding into L^1.*

This conjecture arose from the work of J. Lee and A. Naor, in which it is shown that the nonexistence of such an embedding provides a natural counter-example to the Goemans-Linial conjecture of theoretical computer science; S. Khot and N. Vishnoi gave a first such counterexample. Very roughly, the point is that in some instances, questions in algorithm design, such as the sparsest cut problem, could be solved if it were possible to embed a certain class of finite metric spaces (those with metrics of negative type) into ℓ^1 with universally bounded bi-Lipschitz distortion, i.e., distortion independent of the particular metric and the cardinality.

3.6 Exercises

Exercise 3.6.1. *Let $\mathfrak{g}$ be a Lie algebra with a step s stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$. Denote by $\mathfrak{g}^k$ the k-th element in the lower central series. Show that*

$$\mathfrak{g}^{(k)} = V_k \oplus \cdots \oplus V_s.$$

Exercise 3.6.2. *Show that if a Lie algebra $\mathfrak{g}$ has a step s stratification, then $\mathfrak{g}$ is nilpotent of step s, thus the assumption of a Carnot group being nilpotent is superfluous.*

Exercise 3.6.3. *Let $\mathfrak{h}$ be the Heisenberg Lie algebra generated by the vectors X, Y, and Z with only non-trivial relation $[X, Y] = Z$. Show that the decomposition*

$$\mathfrak{h} = \operatorname{span}\{X, Y\} \oplus \operatorname{span}\{Z\}$$

is a step 2 stratification.

Exercise 3.6.4. *Let $g := \mathbb{R} \times \mathfrak{h}$ be the (commutative) product of $\mathbb{R}$ with the (above) Heisenberg Lie algebra $\mathfrak{h}$. Show that*

$$g = (\mathbb{R} \times \operatorname{span}\{X, Y\}) \oplus (\{0\} \times \operatorname{span}\{Z\})$$

is a step 2 stratification with center $\mathbb{R} \times \operatorname{span}\{Z\}$ which is stricly bigger than V_2.

Exercise 3.6.5. *Show that if a Lie algebra $\mathfrak{g}$ has a step s stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$, then*

1. *V_s is contained in the center of $\mathfrak{g}$;*

2. *$V_k \oplus \cdots \oplus V_s$ is normal in $\mathfrak{g}$; $(V_k \oplus \cdots \oplus V_s)/(V_{k+1} \oplus \cdots \oplus V_s)$ is contained in the center of $(V_1 \oplus \cdots \oplus V_s)/(V_{k+1} \oplus \cdots \oplus V_s)$.*

Exercise 3.6.6. *Show that if G is a Carnot group and Δ is the left-invariant distribution with $\Delta_e = V_1$, then $(\Delta^{[j]})_e = V_1 \oplus \cdots \oplus V_j$.*

Exercise 3.6.7. *Show that if G is a Carnot group and Δ is the left-invariant distribution with $\Delta_e = V_1$, then the three definitions* (2.3.3), (2.6.5), *and* (3.2.2) *of Q coincide.*

Exercise 3.6.8. *Use the BCH formula to show* (3.2.4).

Exercise 3.6.9. *Uses the definitions to prove* (3.2.6).

Exercise 3.6.10. *Show that is any Carnot groups there is a (strong) Malcev basis*

Exercise 3.6.11. *Prove that, if M is a Riemannian manifold, then the Carnot group structure that any T_pM inherits is Abelian.*

Exercise 3.6.12. *Prove that, if M is a contact* 3*-manifold, then the Carnot group structure that any T_pM inherits is the Heisenberg algebra.*

Exercise 3.6.13. *Prove that, if G is a Carnot group, then the Carnot group structure that any T_pG inherits is the Lie algebra $Lie(G)$ itself.*

Exercise 3.6.14. *Give an example of Lie group G with a left-invariant bracket-generating distribution such that Carnot group structure that T_eG inherits is NOT isomorphic to the Lie algebra $Lie(G)$.*

Chapter 4

Analysis on nilpotent groups

4.1 A deeper study of Carnot groups

4.1.1 A good basis for a Carnot group

Let $\mathbb{G}$ be a Carnot group with stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$. We want to construct a basis for $\mathfrak{g}$ that is structured with respect to the stratification, is a Malcev basis, and each element of the basis that is not in V_1, is the bracket of two vectors of such a basis.

Start by picking a basis $X_1, \ldots, X_m$ of V_1. Then consider all brackets $[X_i, X_j]$, for $i, j = 1, \ldots, m$. Since $[V_1, V_1] = V_2$, we can find among such brackets a basis for V_2, cf. Exercise 4.6.1. Pick some such basis and call the elements $X_{m+1}, \ldots, X_{m_2}$. Iterate the method: extract a basis $X_{m_2+1}, \ldots, X_{m_3}$ of V_3 from the set $[X_i, X_j]$, for $i = 1, \ldots, m, j = m+1, \ldots, m_2$. And so on. In such a way we constructed a basis $X_1, \ldots, X_n$ of $\mathfrak{g}$ such that

1. $X_{m_{j-1}+1}, \ldots, X_{m_j}$ is a basis of V_j,

2. For any $i = m+1, \ldots, n$, there exist d_i, l_i, and k_i such that $X_i \in V_{d_i}$, $X_{l_i} \in V_1$, $X_{k_i} \in V_{d_i - 1}$, and
$$X_i = [X_{l_i}, X_{k_i}]. \tag{4.1.1}$$

3. The order-reversed basis $X_n, \ldots, X_1$ is a (strong) Malcev basis; in other words,
$$[\mathfrak{g}, \operatorname{span}\{X_k, \ldots, X_n\}] \subseteq \operatorname{span}\{X_{k+1}, \ldots, X_n\}.$$

We would suggest the terminology '*Carnot basis*' for such basis satisfying the above three conditions. The reader should notice that the above property 1 implies the property 3. See Exercise 4.6.2.

To describe a Carnot algebra we prefer to give a Carnot basis as a hierarchical diagram as

$$\begin{tikzcd}
V_1 : & X \arrow[dr] & & Y \arrow[dl] & \text{for the 3D Heisenberg algebra,} \\
V_2 : & & Z & &
\end{tikzcd}$$

$$\begin{tikzcd}
V_1 : & X \arrow[dr] \arrow[dd] & & Y \arrow[dl] & \text{for the Engel algebra,} \\
V_2 : & & Z \arrow[dl] & & \\
V_3 : & W & & &
\end{tikzcd}$$

$$\begin{tikzcd}
V_1 : & X \arrow[dr] \arrow[dd] & & Y \arrow[dl] \arrow[dd] & \text{for the rank 2 and step 3 algebra.} \\
V_2 : & & Z \arrow[dl] \arrow[dr] & & \\
V_3 : & U & & V. &
\end{tikzcd}$$

The j-th line in the diagram list the vectors that span V_j. The black lines express the non-trivial brackets. However, one should notice that in the algebra structure might be more relations than just those in (4.1.1). (Give an example!)

4.1.2 Local-to-global using dilations, and canonical coordinates

Since any Carnot group is nilpotent and simply connected, the map $\exp \mathfrak{g} \to \mathbb{G}$ is a global diffeomorphism, cf. Theorem 3.1.13. Therefore the exponential coordinates are global (and one-to-one) coordinates. As a consequence, the dilations $\delta_\lambda : \mathbb{G} \to \mathbb{G}$ are well-defined. From them one has that such self-similar homomorphisms extend properties that hold in a neighborhood of the identity to the whole of $\mathbb{G}$. As an example, let us show the fact that Malcev coordinates maps are injective and surjective.

Proposition 4.1.2. *On every Carnot group, Malcev coordinates exist.*

Proof. The fact that a Malcev basis $X_1, \dots, X_n$ exist was shown in the previous subsection. Now consider the coordinate map

$$\Psi : (s_1 \dots, s_n) \to \exp(s_1 X_1) \cdots \exp(s_n X_n).$$

Obviously

$$(d\Psi)_0 \partial_j = \frac{d}{ds_j} \exp(s_j X_j)\Big|_{s_j=0} X_j,$$

so $(d\Psi)_0$ is an invertible $n \times n$ matrix. Thus Ψ is open at zero, i.e., $\Psi(\mathbb{R}^n)$ is a neighborhood of the identity e. Let us show that $\Psi(\mathbb{R}^n) = \mathbb{G}$. Take $p \in \mathbb{G}$. Then there exists some $\lambda \in \mathbb{R}$ and some $\mathbf{s} \in \mathbb{R}^n$ such that

$$\delta_\lambda^{-1}(p) = \Psi(\mathbf{s}).$$

Let $\tilde{\mathbf{s}} = \delta_\lambda(\mathbf{s})$. Then, since δ_λ on $\mathbb{G}$ is a group homomorphism, we have

$$\begin{aligned}
\Psi(\tilde{\mathbf{s}}) &= \exp(\delta_\lambda(s_1 X_1)) \cdots \exp(\delta_\lambda(s_n X_n)) \\
&= \delta_\lambda(\exp(s_1 X_1)) \cdots \delta_\lambda(\exp(s_n X_n)) \\
&= \delta_\lambda(\exp(s_1 X_1) \cdots \exp(s_n X_n)) \\
&= \delta_\lambda \Psi(\mathbf{s}) \\
&= p.
\end{aligned}$$

Let us show injectivity. Since $(d\Psi)_0$ is an invertible $n \times n$ matrix, then by the Inverse Function Theorem there is a neighborhood U on which Ψ is injective. Assume now that there are $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^n$ such that

$$\Psi(\mathbf{s}_1) = \Psi(\mathbf{s}_2).$$

Then, for $\lambda \in \mathbb{R}$ small enough, we have $\delta_\lambda(\mathbf{s}_1), \delta_\lambda(\mathbf{s}_2) \in U$. By the above calculation, we have that

$$\Psi(\delta_\lambda(\mathbf{s}_1)) = \Psi(\delta_\lambda(\mathbf{s}_2)).$$

But, since Ψ is injective on U, we have $\delta_\lambda(\mathbf{s}_1) = \delta_\lambda(\mathbf{s}_2)$, and therefore $\mathbf{s}_1 = \mathbf{s}_2$. $\square$

4.1.3 A direct, effective proof of Chow's Theorem

We will give now an explicit construction of an horizontal path connecting an arbitrary point p in a Carnot group to the origin e. The reader should remind the elementary fact, cf. Theorem 3.1.6, that the curve pe^{tX} is the integral curve of X starting at p.

The brackets as products of exponentials

The philosophy behind the following discussion is that to go in a direction given as a bracket of two vector fields one can go along a quadrilateral constructed using the flows of the two vector fields. We will give a generalization of the following formula with which the reader should be already familiar:

$$[X,Y] = \frac{\mathrm{d}^2}{2\mathrm{d}^2 t} e^{-tY} \circ e^{-tX} \circ e^{tY} \circ e^{tX}\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} e^{-\sqrt{t}Y} \circ e^{-\sqrt{t}X} \circ e^{\sqrt{t}Y} \circ e^{\sqrt{t}X}\Big|_{t=0}.$$

In the above formula, e^{tX} denotes the flow map of a general vector field on a manifold. So for left-invariant vector fields in a Lie group we have

$$e^{tX}(p) = pe^{tX}.$$

Thus the order might seems reversed.

For $X, Y \in \mathfrak{g}$ and $t \in \mathbb{R}$ define

$$P_t(X,Y) := e^{tX} e^{tY} e^{-tX} e^{-tY}.$$

Using twice the BCH formula one has that, for $t \to 0$,

$$P_t(X,Y) = e^{t^2[X,Y]+o(t^2)}.$$

Suppose we have defined by induction the function $P_t(X_1, \ldots, X_k)$, for $k \geq 2$, define then

$$P_t(X_1, \ldots, X_{k+1}) := P_t(X_1, \ldots, X_k) e^{tX_{k+1}} (P_t(X_1, \ldots, X_k))^{-1} e^{-tX_{k+1}}.$$

By induction we shall show that, as $t \to 0$,

$$P_t(X_1, \ldots, X_k) = e^{t^k[\ldots[[X_1,X_2],X_3],\ldots,X_k]+o(t^k)}. \tag{4.1.3}$$

The case $k = 2$ has been already mentioned above, and its proof is similar to the induction step. Assume it true for an arbitrary k. Call $\omega(t)$ the $o(t^k)$ function such that $P_t(X_1, \ldots, X_k) = e^{t^k[\ldots[[X_1,X_2],X_3],\ldots,X_k]+\omega(t)}$. Then we have, by the BCH formula,

$$\begin{aligned}
P_t(X_1, \ldots, X_{k+1}) &= P_t(X_1, \ldots, X_k) e^{tX_{k+1}} \left(P_t(X_1, \ldots, X_k)\right)^{-1} e^{-tX_{k+1}} \\
&= e^{t^k[\ldots[X_1,X_2],\ldots,X_k]+\omega(t)} e^{tX_{k+1}} \left(e^{t^k[\ldots[X_2,X_1],\ldots,X_k]+\omega(t)}\right)^{-1} e^{-tX_{k+1}} \\
&= e^{t^k[\ldots[X_1,X_2],\ldots,X_k]+\omega(t)} e^{tX_{k+1}} e^{-t^k[\ldots[X_2,X_1],\ldots,X_k]-\omega(t)} e^{-tX_{k+1}} \\
&= e^{\left(tX_{k+1}+t^k[\ldots[X_1,X_2],\ldots,X_k]+\omega(t)+\frac{1}{2}t^{k+1}[\ldots[X_1,X_2],\ldots,X_{k+1}]+o(t^{k+1})\right)} \cdot \\
&\qquad e^{\left(-tX_{k+1}-t^k[\ldots[X_1,X_2],\ldots,X_k]-\omega(t)+\frac{1}{2}t^{k+1}[\ldots[X_1,X_2],\ldots,X_{k+1}]+o(t^{k+1})\right)} \\
&= e^{t^k[\ldots[[X_1,X_2],X_3],\ldots,X_{k+1}]+o(t^{k+1})}.
\end{aligned}$$

One should note that each P_t is in fact a product of element of the form $e^{\pm tX_i}$. Thus the following properties are immediate:

$$P_{\lambda t}(X_1, \ldots, X_k) = P_t(\lambda X_1, \ldots, \lambda X_k), \tag{4.1.4}$$

$$\delta_\lambda P_t(X_1,\dots,X_k) = P_t(\delta_\lambda X_1,\dots,\delta_\lambda X_k). \tag{4.1.5}$$

We construct now a map that will help in constructing horizontal paths. Consider a Carnot basis $X_1,\dots,X_n$, so in particular property (4.1.1) holds. Iterating such property, we have that each element X_j of the basis is such that

$$X_j = [\dots[[X_{j,1},X_{j,2}],X_{j,3}],\dots,X_{j,d_j}],$$

where the basis elements $X_{j,1},\dots,X_{j,d_j}$ are in V_1, and d_j is such that $X_j \in V_{d_j}$, in other words, it is the degree of X_j.

For each j, we consider the expression

$$P^{(j)}(t) := P_t(X_{j,1},\dots,X_{j,d_j}).$$

In the following we will use the notation $t^\alpha = \operatorname{sgn}(t)|t|^\alpha$, so for example we have $\sqrt{-4} = -2$. We finally define the map

$$E(\mathbf{t}) := P^{(1)}(\sqrt[d_1]{t_1})\cdots P^{(n)}(\sqrt[d_n]{t_n}).$$

E.g., for the standard basis in the Heisenberg group we get:

$$E(\mathbf{t}) = e^{t_1X}e^{t_2Y}e^{\sqrt{t_3}X}e^{\sqrt{t_3}Y}e^{-\sqrt{t_3}X}e^{-\sqrt{t_3}Y}.$$

For the standard basis in the Engel group we get:

$$\begin{aligned} E(\mathbf{t}) \quad = \quad & e^{t_2X}\, e^{t_2Y}\, e^{\sqrt{t_3}X}e^{\sqrt{t_3}Y}e^{-\sqrt{t_3}X}e^{-\sqrt{t_3}Y} \\ & e^{\sqrt[3]{t_4}X}\, e^{\sqrt[3]{t_4}Y}e^{-\sqrt[3]{t_4}X}e^{-\sqrt[3]{t_4}Y}e^{\sqrt[3]{t_4}X}e^{\sqrt[3]{t_4}Y}e^{\sqrt[3]{t_4}X}e^{-\sqrt[3]{t_4}Y}e^{-\sqrt[3]{t_4}X}\, e^{-\sqrt[3]{t_4}X}. \end{aligned}$$

We will show in order that such a map E satisfies the following three properties.

Proposition 4.1.6. *Let E be the map defined above.*

1. *$E : \mathbb{R}^n \to \mathbb{G}$ is open at $\mathbf{0}$.*

2. *E is surjective.*

3. *E gives a natural horizontal path from $\mathbf{0}$ to $E(\mathbf{t})$.*

The second property follows easily from the first one using dilations. The third is also very elementary since flows of left-invariant vector fields are right multiplications by exponentials. The first is a consequence of the interpretation of the bracket as product of exponential.

Proof of Property 1 of Proposition 4.1.6. We just need to show that $(dE)_0$ is a non-singular matrix. From how E has been defined and from (4.1.3), we have

$$\begin{aligned}
(dE)_0\partial_j &= \frac{d}{dt_j}E(\mathbf{t})\Big|_{\mathbf{t}=0} \\
&= \frac{d}{dt_j}P^{(j)}(\sqrt[d_j]{t_j})\Big|_{t_j=0} \\
&= \frac{d}{dt}P_{\sqrt[d_j]{t}}(X_{j,1},\ldots,X_{j,d_j})\Big|_{t=0} \\
&= \frac{d}{dt}e^{t[\ldots[[X_{j,1},X_{j,2}],X_{j,3}],\ldots,X_{j,d_j}]+o(t)}\Big|_{t=0} \\
&= \frac{d}{dt}e^{tX_j+o(t)}\Big|_{t=0} \\
&= X_j.
\end{aligned}$$

In other words, $(dE)_0$ sends the basis $\partial_1,\ldots,\partial_n$ to the basis $X_1,\ldots,X_n$. Property 1 follows from the Inverse Function Theorem.

Proof of Property 2 of Proposition 4.1.6. By Property 1, the set $E(\mathbb{R}^n)$ is a neighborhood of e. On the other hand for each fixed point $q \in \mathbb{G}$, the dilations δ_λ of the Carnot group have the property that $\lim_{\lambda\to 0}\delta_\lambda(q) = e$. From these two facts we have that, for each $p \in \mathbb{G}$, there are $\lambda \in \mathbb{R}$ and $\mathbf{t} \in \mathbb{R}^n$ such that

$$\delta_\lambda(E(\mathbf{t})) = p.$$

Now, let $\tilde{\mathbf{t}} = \delta_\lambda(\mathbf{t})$, i.e., $\tilde{t}_j = \lambda^{d_j}t_j$. First by the properties (4.1.4) and (4.1.5) on P_t, and the fact that $X_{j,1},\ldots,X_{j,d_j}$ are in V_1, one has

$$\begin{aligned}
P^{(j)}(\sqrt[d_j]{\tilde{t}_j}) &= P^{(j)}(\sqrt[d_j]{\lambda^{d_j}t_j}) \\
&= P^{(j)}(\lambda\sqrt[d_j]{t_j}) \\
&= P_{\lambda\sqrt[d_j]{t_j}}(X_{j,1},\ldots,X_{j,d_j}) \\
&= P_{\sqrt[d_j]{t_j}}(\lambda X_{j,1},\ldots,\lambda X_{j,d_j}) \\
&= P_{\sqrt[d_j]{t_j}}(\delta_\lambda(X_{j,1}),\ldots,\delta_\lambda(X_{j,d_j})) \\
&= \delta_\lambda\left(P_{\sqrt[d_j]{t_j}}(X_{j,1},\ldots,X_{j,d_j})\right) \\
&= \delta_\lambda P^{(j)}(\sqrt[d_j]{t_j}).
\end{aligned}$$

Then, since δ_λ on $\mathbb{G}$ is a group homomorphism, one get

$$\begin{aligned} E(\tilde{\mathbf{t}}) &= P^{(1)}(\sqrt[d_1]{\tilde{t}_1})\cdots P^{(n)}(\sqrt[d_n]{\tilde{t}_n}) \\ &= \delta_\lambda(P^{(1)}(\sqrt[d_1]{t_1}))\cdots\delta_\lambda(P^{(n)}(\sqrt[d_n]{t_n})) \\ &= \delta_\lambda\left(P^{(1)}(\sqrt[d_1]{t_1})\cdots P^{(n)}(\sqrt[d_n]{t_n})\right) \\ &= \delta_\lambda E(\mathbf{t}) \\ &= p. \end{aligned}$$

Thus $E(\mathbb{R}^n)$ is in fact the whole of $\mathbb{G}$, i.e., E is surjective.

Proof of Property 3 of Proposition 4.1.6. Recall, cf. Theorem 3.1.6, that the flow lines of a left-invariant vector field X are the curves ge^{tX}, fixed $g \in \mathbb{G}$ and varying $t \in \mathbb{R}$. Now, since P_t is a product of exponentials, then E is too. More explicitly, fixed $\mathbf{t} \in \mathbb{R}^n$, we have

$$E(\mathbf{t}) = \exp(\xi_1 t_{\gamma_1}^{\alpha_1} X_{\beta_1})\cdots\exp(\xi_N t_{\gamma_N}^{\alpha_N} X_{\beta_N}),$$

for $\xi_i \in \{1, -1\}$, $\alpha_i^{-1} \in \mathbb{N}$, $\beta_i \in \{1, \dots, m\}$, $\gamma_i \in \{1, \dots n\}$, and $N \in \mathbb{N}$. Now it is enough to observe that, fixed K, the point

$$g := \exp(\xi_1 t_{\gamma_1}^{\alpha_1} X_{\beta_1})\cdots\exp(\xi_K t_{\gamma_K}^{\alpha_K} X_{\beta_K})$$

can be connected to the point

$$\exp(\xi_1 t_{\gamma_1}^{\alpha_1} X_{\beta_1})\cdots\exp(\xi_K t_{\gamma_K}^{\alpha_K} X_{\beta_K})\exp(\xi_{K+1} t_{\gamma_{K+1}}^{\alpha_{K+1}} X_{\beta_{K+1}})$$

by the path

$$g\exp(\xi_{K+1} s X_{\beta_{K+1}}), \qquad \text{for } s \in [0, |t_{\gamma_{K+1}}^{\alpha_{K+1}}|],$$

which is tangent to $\pm X_{\beta_{K+1}}$, thus horizontal. □

Corollary 4.1.7 (Chow's Theorem for Carnot groups)**.** *Any point $p \in \mathbb{G}$ in a Carnot group can be joined to the identity e by a horizontal path. Moreover, the CC-distance induces the manifold topology.*

Proof. Property 2 and 3 of Proposition 4.1.6 give the existence of a path from e to any given point p. Thus $d_{CC}(e,p) < \infty$, for all $p \in \mathbb{G}$. By left invariance of d_{CC} we have $d_{CC}(p,q) < \infty$, for all $p, q \in \mathbb{G}$.

Since E is in fact open at $\mathbf{0}$, by Property 1 of Proposition 4.1.6, then points close to the origin can be connected to the origin by short horizontal curves. □

4.1.4 A proof of the Ball-Box Theorem

Let us still denote by d_j the degree of the j-th vector X_j in a Carnot basis, i.e., we have $X_j \in V_{d_j}$. The *box* with respect to a fixed Carnot basis $X_1, \ldots, X_n$ is defined as

$$\mathrm{Box}(r) := \{(t_1, \ldots, t_n) \in \mathbb{R}^n \ : \ |t_j| \leq r^{d_j}\}.$$

Consider the exponential coordinate map $\Phi(\mathbf{t}) = \exp(\sum_j t_j X_j)$. Since the $\exp : \mathfrak{g} \to \mathbb{G}$ is a diffeomorphism, we have that

$$\Phi(\mathrm{Box}(1))$$

is a (bounded) neighborhood of $e \in \mathbb{G}$. Thus, since by Chow's Theorem 4.1.7 the CC distance induces the standard topology, there exists a constant C such that

$$B(e, 1/C) \subseteq \Phi\left(\mathrm{Box}(1)\right) \subseteq B(e, C), \tag{4.1.8}$$

where $B(e,r)$ is the CC-ball of center the origin e and radius r. Recall that

$$\delta_\lambda\left(B(e,r)\right) = B(e, \lambda r).$$

On the other hand,

$$\begin{aligned}
\delta_\lambda\left(\Phi\left(\mathrm{Box}(1)\right)\right) &= \delta_\lambda\left(\{\exp(\sum_j t_j X_j) \ : \ |t_j| \leq 1\}\right) \\
&= \{\exp(\sum_j \lambda^{d_j} t_j X_j) \ : \ |t_j| \leq 1\} \\
&= \{\exp(\sum_j \tilde{t}_j X_j) \ : \ |\tilde{t}_j| \leq \lambda^{d_j}\} \\
&= \Phi\left(\mathrm{Box}(\lambda)\right).
\end{aligned}$$

In general,

$$\delta_\lambda\left(\Phi\left(\mathrm{Box}(r)\right)\right) = \Phi\left(\mathrm{Box}(\lambda r)\right).$$

Thus, (4.1.8) implies

$$\Phi\left(\mathrm{Box}(C^{-1}r)\right) \subseteq B(e,r) \subseteq \Phi\left(\mathrm{Box}(Cr)\right). \tag{4.1.9}$$

Theorem 4.1.10 (Ball-Box Theorem for Carnot groups)**.** *Let $\mathbb{G}$ be a Carnot group. Fix a Carnot basis. Then there exists a constant $C > 1$ such that, for all $p \in \mathbb{G}$, if Φ_p denotes the exponential coordinate map from p with respect to the Carnot basis, then*

$$\Phi_p\left(\mathrm{Box}(C^{-1}r)\right) \subseteq B(p,r) \subseteq \Phi_p\left(\mathrm{Box}(Cr)\right), \tag{4.1.11}$$

for all $r > 0$.

Proof. Notice that

$$\Phi_p(\mathbf{t}) = p \cdot \exp(\sum_j t_j X_j) = L_p\left(\Phi(\mathbf{t})\right).$$

Now, since d_{CC} is left-invariant, then (4.1.9) implies (4.1.11). $\square$

4.1.5 Haar, Lebesgue and Hausdorff measures.

We shall denote by $\mathrm{vol}_{\mathbb{G}}$ the volume form and, at the same time, the right-invariant Haar measure.

Carnot groups are nilpotent and so unimodular, therefore the right and left Haar measures coincide, up to constant multiples. We fix one of them and denote it by $\mathrm{vol}_{\mathbb{G}}$.

We shall denote by $\mathscr{H}^k$ (resp. $\mathscr{S}^k$) the Hausdorff (resp. spherical Hausdorff) k-dimensional measure; these measures depend on the distance, and, unless otherwise stated, to build them we will use the Carnot-Carathéodory distance in $\mathbb{G}$ and the Euclidean distance in Euclidean spaces.

Using the left translation and scaling invariance of the Carnot-Carathéodory distance one can easily check that the Haar measures of $\mathbb{G}$ are a constant multiple of the spherical Hausdorff measure $\mathscr{S}^Q$ and of $\mathscr{H}^Q$. In exponential coordinates, all these measures are a constant multiple of the Lebesgue measure $\mathscr{L}^n$ in $\mathbb{R}^n$, namely

$$\mathrm{vol}_G\big(\{\exp(\sum_{i=1}^n x_i X_i):\ (x_1,\dots,x_n)\in A\}\big) = c\mathscr{L}^n(A) \qquad \text{for all Borel sets } A \subseteq \mathbb{R}^n$$

for some constant c. Using this fact, one can easily prove that

$$\mathrm{vol}_{\mathbb{G}}(\delta_\lambda(A)) = \lambda^Q\,\mathrm{vol}_{\mathbb{G}}(A) \tag{4.1.12}$$

for all Borel sets $A \subseteq \mathbb{G}$.

4.2 Regularity problems

We will discuss the following issues:

- smoothness of geodesic curves;
- smoothness of metric spheres;
- smoothness (and existence) of minimal surfaces;
- smoothness (and existence) of solution of the isoperimetric problem.

Comments regarding geodesics

1. The existence is ensured by Ascoli-Arzelà Theorem, as a priori just Lipschitz curves, so differentiable almost everywhere.
2. People expect that when $(M, \Delta, \langle\cdot,\cdot\rangle)$ is a sub-Riemannian manifold, then any geodesic is C^1, or, in fact, C^∞. The question is still open.
3. People expect that when $\|\cdot\|$ is a norm coming from a polytope, i.e., the unit ball of $\|\cdot\|$ is the convex hull of finitely many points, then there exists a constant $N \in \mathbb{N}$ such that each pair of points can be connected with a geodesic made of N smooth pieces. The question is still open.
4. The query cannot be solved using the standard arguments from geometric analysis (e.g., Calculus of Variation or differential geometry) as in Riemannian geometry.

Comments regarding metric spheres

1. In Carnot groups, metric spheres are topological spheres. (In general, the conjecture is that small metric spheres are topological spheres.)
2. In the Heisenberg geometry, spheres are not smooth at the pole. See the picture of the section of the ball.
3. The expectation is that small metric spheres (at least in Carnot groups) should be piecewise smooth.
4. The regularity of geodesics is linked (at least philosophically) to the regularity of metric spheres.

Comments regarding minimal surfaces and isoperimetric solutions

1. They do exists in an extended sense.
2. Regularity is a tricky issue.

4.2.1 Common general philosophical strategy for regularity

Step 1 Consider the geometric objects as special elements inside a wider class of analytical objects.

Step 2 Prove that such analytical objects are in fact 'rectifiable', e.g., 'piece-wise Lipschitz'. (Here there will be an issue since Carnot groups are purely unrectifiable.)

Step 3 Rectifiability should be first improved as low (e.g., C^1) regularity, for example in the case of minimal objects.

Step 4 Minimal C^1 (or C^2) objects are in fact C^∞, or even analytic.

4.3 Generalized hyper-surfaces: sets with finite perimeter

Both metric spheres and $(n-1)$-dimensional minimal surfaces inside an n-dimensional Carnot group have codimension 1. We can see them as boundary of an n-dimensional domain Ω. We then think about studying Ω instead $\partial\Omega$. The idea is to consider the characteristic function χ_Ω of Ω:

$$\chi_\Omega(x) = 1 \text{ if } x \in \Omega, \chi_\Omega(x) = 0 \text{ if } x \notin \Omega.$$

We consider the wide class of all measurable sets Ω, in other words, we have $\chi_\Omega \in L^1_{loc}$.

Which are the good χ_Ω? Clearly, even a request of continuity is too strong. The feeling is that if Ω is a hyper-space, then χ_Ω should be good. As an toy example, let us consider the I-don't-remember-the-name function, i.e. $\chi_{\mathbb{R}_{>0}}$. A nice property of such a function is that its derivative exists in the generalized sense, it is the delta measure δ_0.

We arrive at the conclusion that "our good sets are those whose characteristic functions have measures as generalized derivatives." We should explain in the following what is this generalized derivative.

4.3.1 A review of divergence and distributions

Let M be any smooth differentiable manifold with topological dimension n, endowed with an n-differential volume form vol_M. For example, vol_M could be a Riemannian volume form; however, eventually, M will be a Lie group $\mathbb{G}$, and vol_M a right Haar measure.

We use the volume form to define the divergence as follows:

Definition 4.3.1. *For any vector field $X \in \Gamma(M)$ define the function* $\operatorname{div} X : M \to \mathbb{R}$ *implicitly as*

$$\int_M Xu\, d\,\mathrm{vol}_M = -\int_M u \operatorname{div} X\, d\,\mathrm{vol}_M \qquad \forall u \in C_c^\infty(M). \tag{4.3.2}$$

We say that X is divergence-free if $\operatorname{div} X \equiv 0$.

For example the vector fields $\dfrac{\partial}{\partial_j}$ in $\mathbb{R}^n$ are divergence-free, because of the Fundamental Theorem of Calculus and the fact that the test functions have compact support.

When (M, g) is a Riemannian manifold and vol_M is the volume form induced by g, then an explicit expression of this differential operator can be obtained in terms of the components of X, and (4.3.2) corresponds to the divergence theorem on manifolds. We won't need either a Riemannian structure or an explicit expression of $\operatorname{div} X$ in the sequel, and for this reason we have chosen a definition based on (4.3.2): this emphasizes the dependence of $\operatorname{div} X$ on vol_M only.

Note that by Leibniz rule $X(uv) = uXv + vXu$, integrating over the manifold when X is a divergence-free vector field, one obtains

$$\int_M uXv\, d\,\mathrm{vol}_M = -\int_M vXu\, d\,\mathrm{vol}_M \qquad \forall u,\, v \in C_c^\infty(M). \tag{4.3.3}$$

This last identity motivates the following classical definition.

Definition 4.3.4 (X-distributional derivative). *Let $u \in L^1_{\mathrm{loc}}(M)$ and let $X \in \Gamma(TM)$ be divergence-free. The generalized derivative of u in the direction of X is the operator $Xu \in (C_c^\infty(M))^*$ defined as*

$$\langle Xu, v\rangle := -\int_M uXv\, d\,\mathrm{vol}_M, \qquad v \in C_c^\infty(M).$$

If $f \in L^1_{\mathrm{loc}}(M)$, we write $Xu = f$ if $\langle Xu, v\rangle = \int_M vf\, d\,\mathrm{vol}_M$ for all $v \in C_c^\infty(M)$. Analogously, if μ is a Radon measure in M, we write $Xu = \mu$ if $\langle Xu, v\rangle = \int_M v\, d\mu$ for all $v \in C_c^\infty(M)$.

Since X is divergence-free and so (4.3.3) holds (it is still valid when $u \in C^1(M)$), the distributional definition of Xu is equivalent to the classical one whenever $u \in C^1(M)$.

Proposition 4.3.5. **(i)** *Let $\mathbb{G}$ be a nilpotent Lie group, and let $\mathrm{vol}_{\mathbb{G}}$ be a right Haar measure. Then each left invariant vector field is divergence-free.*

(ii) *More generally, for any manifold M, any volume form vol_M), and any $X \in \Gamma(M)$, one has that if the flows of X are vol_M-preserving, then $\operatorname{div} X \equiv 0$.*

Proof. The first assertion is consequence of the second one, since, as we saw, flows of left invariant vector fields are right translations, $g \mapsto ge^{tX}$. Regarding (ii), let $\Phi_X^t(\cdot)$ be the flow of X at time t. Thus we know that, for any t, we have

$$\left(\Phi_X^t\right)_\# \mathrm{vol}_M = \mathrm{vol}_M\,.$$

Therefore, for any test function u, $\int_M u \circ \Phi_X^t \, d\,\mathrm{vol}_M = \int_M u \, d\,\mathrm{vol}_M$. Such independence of t implies that

$$\begin{aligned}
-\int_M u \operatorname{div} X \, d\,\mathrm{vol}_M &= \int_M Xu \, d\,\mathrm{vol}_M \\
&= \int_M (Xu) \circ \Phi_X^t \, d\,\mathrm{vol}_M \\
&= \int_M \frac{d}{dt} \left(u \circ \Phi_X^t \right) \, d\,\mathrm{vol}_M \\
&= \frac{d}{dt} \int_M u \circ \Phi_X^t \, d\,\mathrm{vol}_M \\
&= 0.
\end{aligned}$$

Therefore $\int_M u \operatorname{div} X \, d\,\mathrm{vol}_M = 0$ for all $u \in C_c^1(M)$, and X is divergence-free. $\square$

One can prove the inverse implication: the flows are vol_M-measure preserving if $\operatorname{div} X$ is equal to 0, cf. the proof of Theorem 2.12 in [AKLD08].

4.3.2 Caccioppoli sets: sets of locally finite perimeter

Definition 4.3.6 (Sets of locally finite perimeter)**.** *A Borel set E in a Carnot group, with stratification $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$, is said a* Caccioppoli set *or to have* locally finite perimeter *if, for any left invariant horizontal vector field $X \in V_1$, the distribution $X\chi_E$ is a Radon measure.*

Now that we generalized the object of study, we should first understand how to obtain back our hyper-surfaces.

Pick $X_1, \dots, X_m$ a basis of V_1. We form the $\mathbb{R}^m$-valued Radon measure

$$D\chi_E := (X_1\chi_E, \dots, X_m\chi_E), \tag{4.3.7}$$

and call it the *perimeter vector measure.* One can write

$$D\chi_E = \nu_E |D\chi_E|,$$

where $|D\chi_E|$ is the (positive) measure given by the variation of $D\chi_E$: if A is any Borel set, then

$$|D\chi_E|(A) = \sup_{\pi} \sum_{B \in \pi} \|D\chi_E(B)\|\,,$$

where the supremum is taken over all partitions π of A into a finite number of disjoint measurable subsets. And ν_E is the vector measurable function obtained as

$$\nu_E(x) := \lim_{r \downarrow 0} \frac{D\chi_E(B_r(x))}{|D\chi_E|(B_r(x))},$$

which exists $|D\chi_E|$-almost everywhere.

The terminology is that $|D\chi_E|$ is the *perimeter measure*, and ν_E is the *normal* of the set. Finally,

$$\mathrm{Per}(E) := |D\chi_E|(\mathbb{G}) \tag{4.3.8}$$

is the *perimeter* of E. More generally, if Ω is a Borel set, then $\mathrm{Per}(E, \Omega) := |D\chi_E|(\Omega)$ is the perimeter of E inside Ω.

All such objects depend on the choice of $X_1, \dots, X_m$. The choice of such a basis is in correspondence to the choice of a sub-Riemannian metric on the Carnot group $\mathbb{G}$, for which $X_1, \dots, X_m$ is an orthonormal basis.

Definition 4.3.9 (De Giorgi's reduced boundary)**.** *Let $E \subseteq \mathbb{G}$ be a set of locally finite perimeter. Define the reduced boundary $\mathscr{F}E$ as the set of points $x \in \operatorname{supp}|D\chi_E|$ where:*

(i) the limit defining ν_E exists and

(ii) $|\nu_E(x)| = 1$.

E.g., the reduced boundary of a square on the (Euclidean) plane is formed by its four edges with the four vertices removed.

Why it is better to consider such sets? Because in such class minima always exist.

Theorem 4.3.10 (Compactness [GN96] + Lower semicontinuity for BV functions [FSSC96])**.** *Let $\mathbb{G}$ be a Carnot group and let E_j be a sequence of locally finite perimeter sets such that their perimeters in some Borel set Ω converge to a value $c \in \mathbb{R}$, i.e.,*

$$|D\chi_{E_j}|(\Omega) \to c.$$

Then there exists a locally finite perimeter set F such that, up to passing to a subsequence,

1. *$\chi_{E_j} \to \chi_F$ in $L^1_{loc}(\Omega)$ and*
2. *$|D\chi_F|(\Omega) \le c$.*

4.3.3 Notions of rectificability

In general metric spaces the classical definition of 'good' surfaces goes back at least to Federer (see [Fed69, 3.2.14]). The 'good' surfaces are those that are images of open subsets in Euclidean spaces via Lipschitz maps.

However, there is a problematic fact: in the Heisenberg group there are no Lipschitz embedding of an open set $U \subset \mathbb{R}^2$ into the group. Indeed, differentiability theorems implies that the Heisenberg group is 2-purely unrectifiable, cf. [AK00, Theorem 7.2]. This means that each Lipschitz map $f : U \subset \mathbb{R}^2 \to \mathbb{G}$ is such that $\mathcal{H}^2(f(U)) = 0$. Roughly speaking, since the 3D Heisenberg group has Hausdorff dimension equal to 4, then the metric dimension of a hyper-surface is espected to be $4 - 1 = 3$. But the image by a Lipschitz map of a 2-dimensional Euclidean set has Hausdorff dimension no greater than 2.

There is a second notion (cf. [FSSC03, FSSC01]) of good surfaces which is only valid for hyper-surfaces: being (locally) the zero set of a 'intrinsically' C^1 real-valued function with non-vanishing gradient:

Definition 4.3.11 ($\mathbb{G}$-regular functions and hyper-surfaces)**.** *Let $\mathbb{G}$ be a Carnot group with V_1 as horizontal layer. Let U be an open subset of $\mathbb{G}$ and $f : U \to \mathbb{R}$. We say that f belongs to $C^1_{\mathbb{G}}(U)$ if f and X_f are continuous functions in U, for all $X \in V_1$. We say that $S \subset \mathbb{G}$ is a* $\mathbb{G}$-regular hyper-surface *if for any $p \in S$ there is an neighborhood U of p in $\mathbb{G}$ and there is $f \in C^1_{\mathbb{G}}(U)$ with $(Xf)(q) \neq 0$, for all $q \in U$ and all $X \in V_1 \setminus \{0\}$, such that*

$$S \cap U = f^{-1}(0).$$

Notice that if f is in C^1 then it is clearly in $C^1_{\mathbb{G}}$. However, the hyper-surface $f^{-1}(0)$ is $\mathbb{G}$-regular only if ∇f is never orthogonal to V_1.

Definition 4.3.12 ($\mathbb{G}$-rectifiable hyper-surface)**.** *Let $\mathbb{G}$ be a Carnot group of Hausdorff dimension Q. A set $\Sigma \subset \mathbb{G}$ is said ($(Q-1)$-dimensional)* $\mathbb{G}$-rectifiable *if there exist a countable collection of $\mathbb{G}$-regular hyper-surfaces S_j such that*

$$\mathcal{H}^{Q-1}_{cc}(\Sigma \setminus \cup_j S_j) = 0.$$

The following theorem is due to De Giorgi in the Euclidean setting and to Franchi, Serapioni, and Serra Cassano in Carnot groups of step 2, cf. [DG54, DG55, FSSC03, FSSC01].

Theorem 4.3.13 (Structure of finite perimeter sets)**.** *Let $\mathbb{G}$ be either the Euclidean space or a step-2 Carnot group. If E has locally finite perimeter, then its reduced boundary $\mathscr{F}E$ is $\mathbb{G}$-rectifiable.*

Question 4.3.14. *Is the above theorem true in Carnot groups of arbitrarily step?*

A partial answer to the above question has been obtained in [AKLD08].

4.3.4 Notions of surface measures

We reach the conclusion that the problem of studying hyper-surfaces can be rephrased as the study of characteristic functions χ_E, focusing on their perimeter measures $|D\chi_E|$ and their reduce boundaries $\mathscr{F}E$. The reason for doing so is that perimeters have properties of compactness and lower semicontinuity, cf. Theorem 4.3.10.

For hyper-surfaces then we have that there are two natural notions of measures: $\mathcal{H}_{cc}^{Q-1}$ restricted to the hyper-surface or the perimeter of one of the side domains determined by the hyper-surface. People expect that the two notions should be related. For doing so, one should first prove rectifiability of reduced boundaries, cf. Question 4.3.14.

However, if $S = f^{-1}(0)$ is given as level set of a C^1 function f, the two measures are equal. Indeed, let $E = f^{-1}((-\infty, 0))$, so $\partial E = S$. Then

$$\begin{aligned} \mathrm{Per}(E) &= \\ &= ... \\ &= \mathcal{H}_{cc}^{Q-1} \llcorner \partial E \end{aligned}$$

4.4 Partial regularity results and open questions

4.4.1 Results on geodesics

The following theorem can be found in [Str86], however, in that paper the claim was wrongly stated in more generality. In fact, the proof was valid only for step-2 distributions. The paper has been corrected in [Str89].

Theorem 4.4.1 (Strichartz [Str89]). *If $(M, \Delta, \langle \cdot, \cdot \rangle)$ is a sub-Riemannian manifold of step-2, then each geodesic for the CC-distance is C^∞.*

The following theorem is proved more generally in [LM08], however the assumptions of step ≤ 4 and rank 2 are deeply used.

Theorem 4.4.2 (Leonardi-Monti). *[LM08] If $\mathbb{G}$ is a Carnot group of step ≤ 4 and with 2-dimensional horizontal layer V_1, then each geodesic for the CC-distance is C^∞.*

The next result is proved in these notes.

Proposition 4.4.3. *Let G be a connected, simply connected, and nilpotent Lie group. Let $\Delta \subset \mathfrak{g}$ be a left-invariant sub-bundle such that*

$$\Delta \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}.$$

E.g., G could be a Carnot group. Then, if X is a left-invariant vector field in Δ, then $t \mapsto e^{tX}$ is a (smooth) geodesic with respect to any CC-distance of $(G, \Delta, \|\cdot\|)$, for any left-invariant norm $\|\cdot\|$.

The following theorem should be found in [Bre07].

Theorem 4.4.4 (Breuillard 2007)**.** *Let $\mathbb{G}$ be the 3D Heisenberg group. Let $\|\cdot\|_1$ be the ℓ^1 norm on V_1. Then the geodesics with respect to the CC-distance of $(G, V_1, \|\cdot\|_1)$ are made of at most 4 pieces of horizontal lines, i.e., each geodesic is the concatenation of at most 4 curves of the form $t \mapsto ge^{tX}$, with $g \in \mathbb{G}$ and $X \in V_1$.*

Conjecture 4.4.5 (Regularity conjecture for sub-Reimannian manifolds)**.** *If $(M, \Delta, \langle \cdot, \cdot \rangle)$ is a sub-Riemannian manifold, then each geodesic for the CC-distance is C^∞.*

Conjecture 4.4.6 (Weak regularity conjecture for sub-Reimannian Carnot groups)**.** *If $\mathbb{G}$ is a Carnot group, then each pair of points can be connected by a C^1 geodesic.*

Conjecture 4.4.7 (Regularity Conjecture for sub-Finsler Carnot groups)**.** *If $(\mathbb{G}, V_1, \|\cdot\|_1)$ is a Carnot group where $\|\cdot\|_1$ is the ℓ^1 norm, then there exists a constant K such that each pair of points can be connected by a geodesic that is the concatenation of at most K horizontal lines.*

There are several statements that are true but for possibly a measure-zero collection of distributions. Compare the following result with Theorem 4.4.12.

Theorem 4.4.8 (Chitour-Jean-Trélat [CJT06])**.** *For generic sub-Riemannian structures $(M, \Delta, \langle \cdot, \cdot \rangle)$ of rank greater than or equal to 3, i.e., $\dim \Delta_p \geq 3$, for all $p \in M$, all geodesics for the CC-distance are C^∞.*

4.4.2 Results on metric spheres

Proposition 4.4.9. *If $\mathbb{G}$ is the 3D Heisenberg group, then each metric sphere $\partial B(e, r)$, $r > 0$, is an (Euclidean) Lipschitz manifolds, and there are two points p_N and p_S (the two poles) such that $\partial B(e, r) \setminus \{p_N, p_S\}$ is a C^∞ manifold.*

In the Carnot group setting, one can uses the dilations and the standard proof of the fact that open sets that are star-shaped are topological balls, to prove that metric balls in Carnot groups are topological balls. Moreover, the spheres can be written as graphs using 'inhomogeneous' spherical coordinates with respect to the dilations. Since metric spheres in CC-metrics are closed, one get the following result.

Proposition 4.4.10. *If $\mathbb{G}$ is a Carnot group, then each metric sphere $\partial B(e, r)$, $r > 0$, is topologically a sphere.*

The following theorem should be found in [Bre07].

Theorem 4.4.11 (Breuillard 2007)**.** *Let $\mathbb{G}$ be the 3D Heisenberg group. Let $\|\cdot\|_1$ be the ℓ^1 norm on V_1. Then the metric spheres of the sub-Finsler geometry of $(G, V_1, \|\cdot\|_1)$ are piece-wise analytical sub-variety.*

The work of Agracev and Gauthier [AG01] gives an piece-wise analytic answer in generic cases:

Theorem 4.4.12 (Agrachev-Gauthier)**.** *Generically, small balls in a sub-Riemannian manifold $(M, \Delta, \langle\cdot,\cdot\rangle)$ are sub-analytic if the rank of the distribution is ≥ 3.*

Conjecture 4.4.13. *If $(M, \Delta, \|\cdot\|)$ is any sub-Finsler manifold, then small metric spheres are piece-wise smooth.*

Proposition 4.4.14. *Metric balls in Carnot groups are sets of finite perimeter and metric spheres are $\mathbb{G}$-rectifiable hyper-surfaces.*

4.4.3 Results on the isoperimetric problem

In studying minimal problems for hyper-surfaces inside a Carnot group $\mathbb{G}$ of Hausdorff dimension Q, it is more convenient to minimize the intrinsic perimeter of a class of sets $E \subset \mathbb{G}$ than the $(Q-1)$-dimensional Hausdorff measure of their boundaries.

Theorem 4.4.15 (Existence of isoperimetric sets)**.** *In any Carnot group, there exist solutions of the isoperimetric problem, i.e., sets minimizing the intrinsic perimeter among all measurable sets with prescribed volume measure.*

The above theorem is due, in the Carnot group setting to Leonardi and Rigot in [LR03], and it has been then generalized by Danielli, Garofalo, and Nhieu.

Proposition 4.4.16. *Metric spheres $\partial B(e, r)$, $r > 0$, in the Heisenberg group are not solutions of the isoperimetric problem.*

In [Pan82, Pan83a], Pierre Pansu draw attention on a class of sets which are called today *Pansu spheres*. Denote by $\mathbb{S}_\lambda$ the compact embedded surface of revolution, which is homeomorphic to a sphere, obtained considering a geodesic between two points in the center of the group at distance π/λ and rotating such a curve around the center. Any left traslation of an $\mathbb{S}_\lambda$ is called a Pansu sphere.

Ritoré and Rosales arrived at a characterization of complete, oriented, connected C^2 immersed volume preserving area-stationary surfaces in the 3D Heisenberg group [RR08, Theorems 6.1, 6.8, 6.11], which led to a proof of the Pansu conjecture (cf. [Pan83a, page 172]) for the isoperimetric profile of the Heisenberg group in the C^2-smooth category [RR08, Theorem 7.2].

Theorem 4.4.17 (Ritoré and Rosales [RR08])**.** *In the 3D Heisenberg group, C^2 isoperimetric sets are Pansu spheres.*

Theorem 4.4.18 (Monti-Rickly [MR09])**.** *(Euclidean) convex isoperimetric sets are Pansu spheres.*

4.4.4 Results on minimal surfaces

Let S be a hyper-surface inside a Carnot group $\mathbb{G}$ of Hausdorff dimension Q. The first two natural surface measures on S are the $(Q-1)$-Hausdorff measure $\mathcal{H}^{Q-1}_{cc} \llcorner_S$ or the perimeter measure of one of the side regions determined by S, i.e., $\mathrm{Per}(E)$ with $\partial E = S$, where the perimeter has been defined in (4.3.8). The perimeter measure $\mathrm{Per}(E)$ has a better behavior and, at least when ∂E is a C^2 hyper-surface, it coincides with $\mathcal{H}^{Q-1}_{cc} \llcorner_{\partial E}$

Let us clearify now the terminology of 'minimal surface'.

Definition 4.4.19. *If $\Sigma \subset \mathbb{G}$ is such that for all Σ' such that there exists $R > 0$ such that*

[...] then we say that Σ is globally area-minimizing

Definition 4.4.20 (...)**.** *then we say that Σ is* (locally) area-minimizing

Definition 4.4.21 (...)**.**

$$-\nabla_{\mathbb{G}} \cdot \frac{\nabla_{\mathbb{G}} F}{|\nabla_{\mathbb{G}} F|} \equiv 0, \qquad \text{where } \nabla_{\mathbb{G}} f = (X_1 f, \dots, X_m f), \tag{4.4.22}$$

then we say that Σ has zero mean curvature *or that it is a* solution of the minimal surface equation.

Definition 4.4.23 (...)**.** *then we say that Σ is* area-stationary*.*

With the term 'minimal surface' authors can reefer to any of the 4 above definitions.

Theorem 4.4.24 (Existence of area-minimizing sets [GN96])**.** *In sub-Riemannian manifolds, area-minimizing sets exist.*

Explicitly, let Ω be a bounded open set in a Carnot group $\mathbb{G}$. Let L be a locally finite perimeter set. Then the above theorem guarantees the existence of a locally finite perimeter set E such that

i) $(E\Delta L) \setminus \Omega = \emptyset$, and

ii) $(F\Delta L) \setminus \Omega = \emptyset \implies Per(E \cap \Omega) \leq Per(F \cap \Omega)$.

In other words, the (reduced) boundary of E is the area minimizing (generalized) hyper-surface inside Ω with boundary data L outside Ω.

Cheng, Hwang and Yang [CHY07] have studied the weak solutions of the minimal surface equation for intrinsic graphs in the Heisenberg group and have proven existence and uniqueness results.

Fact: The minimal surface equation is a sub-elliptic PDE: a priori, neither existence, not uniqueness, nor regularity can be deduced.

Theorem 4.4.25 (Non-uniqueness of minimal surfaces [Pau04])**.** *There are loops in the Heisenberg group that admit more than one filling by zero-mean curvature disks.*

N.B. This happens in the Euclidean case too.

The main difference between Euclidean and sub-Riemannian geometry is the existence of low-regular minimal surfaces.

Theorem 4.4.26 (Existence of low-regular area minimizing surfaces [Rit09, CHY07, Pau04])**.** *There are area-minimizing surfaces in the 3D Heisenberg group that are not C^2.*

This is due to the fact that not all area-minimizing surfaces have zero-mean curvature. On the other hand, there are examples of zero-mean curvature surfaces that are not area-minimizing, cf. [DGN08]. Does this happen in Euclidean geometry?

Moreover, the condition of having zero mean curvature is not enough to guarantee that a given surface of class C^2 is area-stationary [RR08].

Theorem 4.4.27 (Regularity of zero mean curvature surfaces [Pau06, CHY09, CCM08])**.** *Let S be a surface in the Heisenberg group that is either C^1 or a Lipschitz intrinsic graphs. If S have zero mean curvature (in an extended sense), then it is smooth.*

Theorem 4.4.28 (Bernstein problem)**.** *In the Euclidean 3D space, any entire minimizing graph $\{(x, y, f(x, y) : x, y \in \mathbb{R}\}$ is a plane.*

One would expect that such a fact would be true for any n-dimensional graph in $\mathbb{R}^{n+1}$, but Bombieri, De Giorgi and Giusti established the surprising result that the Bernstein property fails if $n \leq 8$.

Theorem 4.4.29 (Counterexample in $\mathbb{R}^9$, [BDGG69])**.** *If $n \leq 8$ there exist complete minimal graphs in $\mathbb{R}^{n+1}$ that are not hyper-planes: For $m \geq 4$, a Simons cone, i.e., the set $E \subset \mathbf{R}^4$ defined by ${x_1}^2 + {x_2}^2 + \cdots + {x_m}^2 = x_{m+1}^2 + x_{m+2}^2 + \cdots + {x_{2m}}^2$ is a minimal surface.*

Theorem 4.4.30 (Counterexample in Heisenberg-Garofalo and Pauls)**.** *Let $G \sim \mathbb{R}^3$ be the Heisenberg group. The real analytic surface*

$$S = \{(x, y, t) \in G | y = -x \tan(\tanh(t))\},$$

is an entire graph with zero mean curvature.

4.4.5 More results on regularity

The work of Agracev and Gauthier [AG01] gives an analytic answer in generic cases:

Theorem 4.4.31 (Agrachev-Gauthier)**.** *Generically, the germ at a point q_0 of the function $q \mapsto \rho(q) \overset{\text{def}}{\to}= \operatorname{dist}(q, q_0)$ is subanalytic if the dimension n of the manifold and the dimension k of the distribution satisfy $n \leq (k-1)k + 1$.*

Theorem 4.4.32 (Agrachev-Gauthier)**.** *Generically (and, in fact, on the complement of a set of distributions of infinite codimension), small balls $\{q\colon \rho(q) \leq r\}$ are subanalytic if $k \geq 3$.*

Theorem 4.4.33 (Agrachev-Gauthier)**.** *Generically, the germ of ρ at q_0 is not subanalytic if $n \geq (k-1)\Big(\frac{k^2}{3} + \frac{5k}{6} + 1\Big)$.*

(Monti, 2000, 2003), (Leonardi-Masnou, 2005): There is no direct counterpart of the Brunn-Minkowski inequality in Euclidean space

(Ritor -Rosales, 2005), (Danielli-Garofalo-Nhieu, 2006): The sets bounded by S_λ are isoperimetric regions in restricted classes of sets (C^2 rotationally symmetric and C^1 unions of two graphs over a ball in the xy-plane $t=0$ divided by $t=0$ into two regions of equal volumes)

Bonk-Capogna: flow by mean curvature of a C^2 convex surface which is the union of two radial graphs, converges to S_λ

4.5 Translations and flows

Given $X \in \Gamma(TM)$ we can consider the associated flow, i.e., the solution $\Phi_X : M \times \mathbb{R} \to M$ of the following ODE

$$\begin{cases} \dfrac{d}{dt}\Phi_X(p,t) &= X_{\Phi_X(p,t)} \\ \Phi_X(p,0) &= p. \end{cases} \tag{4.5.1}$$

Notice that the smoothness of X ensures uniqueness, and therefore the semigroup property

$$\Phi_X(x,t+s) = \Phi_X(\Phi_X(x,t),s) \qquad \forall t,\ s \in \mathbb{R},\ \forall x \in M \tag{4.5.2}$$

but not global existence; it is guaranteed, however, for left-invariant vector fields in Lie groups. We obviously have

$$\frac{d}{dt}(u \circ \Phi_X)(p,t) = (Xu)(\Phi_X(p,t)) \qquad \forall u \in C^1(M). \tag{4.5.3}$$

An obvious consequence of this identity is that, for a C^1 function u, $Xu = 0$ implies that u is constant along the flow, i.e. $u \circ \Phi_X(\cdot,t) = u$ for all $t \in \mathbb{R}$. A similar statement holds even for distributional derivatives along vector fields: for simplicity let us state and prove this result for divergence-free vector fields only.

Theorem 4.5.4. *Let $u \in L^1_{\mathrm{loc}}(M)$ be satisfying $Xu = 0$ in the sense of distributions. Then, for all $t \in \mathbb{R}$, $u = u \circ \Phi_X(\cdot,t)$ vol_M-a.e. in M.*

Proof. Let $g \in C^1_c(M)$; we need to show that the map $t \mapsto \int_M gu \circ \Phi_X(\cdot,t)\,d\,\mathrm{vol}_M$ is independent of t. Indeed, the semigroup property (4.5.2), and the fact that X is divergence-free yield

$$\begin{aligned} & \int_M gu \circ \Phi_X(\cdot,t+s)\,d\,\mathrm{vol}_M - \int_M gu \circ \Phi_X(\cdot,t)\,d\,\mathrm{vol}_M \\ = & \int_M ug \circ \Phi_X(\cdot,-t-s)\,d\,\mathrm{vol}_M - \int_M ug \circ \Phi_X(\cdot,-t)\,d\,\mathrm{vol}_M \\ = & \int_M ug \circ \Phi_X(\Phi_X(\cdot,-s),-t)\,d\,\mathrm{vol}_M - \int_M ug \circ \Phi_X(\cdot,-t)\,d\,\mathrm{vol}_M \\ = & -s\int_M uX(g \circ \Phi_X(\cdot,-t))\,d\,\mathrm{vol}_M + o(s) = o(s). \end{aligned}$$

□

Remark 4.5.5. We notice also that the flow is vol_M-measure preserving (i.e. $\mathrm{vol}_M(\Phi_X(\cdot,t)^{-1}(A)) = \mathrm{vol}_M(A)$ for all Borel sets $A \subseteq M$ and $t \in \mathbb{R}$) if and only if $\operatorname{div} X$ is equal to 0. Indeed, if $f \in C^1_c(M)$, the measure preserving property gives that $\int_M f(\Phi_X(x,t))\, d\,\mathrm{vol}_M(x)$ is independent of t. A time differentiation and (4.5.3) then give

$$0 = \int_M \frac{d}{dt} f(\Phi_X(x,t))\, d\,\mathrm{vol}_M(x) = \int_M Xf(\Phi_X(x,t))\, d\,\mathrm{vol}_M(x) = \int_M Xf(y)\, d\,\mathrm{vol}_M(y).$$

Therefore $\int_M f \operatorname{div} X\, d\,\mathrm{vol}_M = 0$ for all $f \in C^1_c(M)$, and X is divergence-free. The proof of the converse implication is similar, and analogous to the one of Theorem 4.5.4.

Let $\mathbb{G}$ be a Lie group with Lie algebra $\mathfrak{g}$. We shall also consider as volume form $\mathrm{vol}_{\mathbb{G}}$ a right-invariant Haar measure.

Let $X \in \mathfrak{g}$ and let us denote, as usual in the theory, by $\exp(tX)$ the flow of X at time t starting from e (that is, $\exp(tX) := \Phi_X(e,t) = \Phi_{tX}(e,1)$); then, the curve $g\exp(tX)$ is the flow starting at g: indeed, since X is left-invariant, setting for simplicity $\gamma(t) := \exp(tX)$ and $\gamma_g(t) := g\gamma(t)$, we have

$$\frac{d}{dt}\gamma_g(t) = \frac{d}{dt}(L_g(\gamma(t))) = (dL_g)_{\gamma(t)}\frac{d}{dt}\gamma(t) = (dL_g)_{\gamma(t)}X = X_{\gamma_g(t)}.$$

This implies that $\Phi_X(\cdot,t) = R_{\exp(tX)}$ and so the flow preserves the right Haar measure, and the left translation preserves the flow lines. By Remark 4.5.5 it follows that all $X \in \mathfrak{g}$ are divergence-free, and Theorem 4.5.4 gives

$$f \circ R_{\exp(tX)} = f \quad \forall t \in \mathbb{R} \qquad \Longleftrightarrow \qquad Xf = 0 \tag{4.5.6}$$

whenever $f \in L^1_{\mathrm{loc}}(\mathbb{G})$.

4.5.1 X-derivative of nice functions and domains

If u is a C^1 function in $\mathbb{R}^n$, then Xu can be calculated as the scalar product between X and the gradient of u:

$$Xu = \langle X, \nabla u\rangle. \tag{4.5.7}$$

Assume that $E \subset \mathbb{R}^n$ is locally the sub-level set of the C^1 function f and that $X \in \Gamma(T\mathbb{R}^n)$ is divergence-free. Then, for any $v \in C^\infty_c(\mathbb{R}^n)$ we can apply the Gauss–Green formula to the vector field vX, whose divergence is Xv, to obtain

$$\int_E Xv\, dx = \int_{\partial E} \langle vX, \nu_E^{eu}\rangle\, d\mathscr{H}^{n-1},$$

where ν_E^{eu} is the unit (Euclidean) outer normal to E. This proves that

$$X\chi_E = -\langle X, \nu_E^{eu}\rangle \mathscr{H}^{n-1} \llcorner_{\partial E}.$$

However, we have an explicit formula for the unit (Euclidean) outer normal to E, it is $\nu_E^{eu}(x) = \nabla f(x)/|\nabla f(x)|$, so, by (4.5.7),

$$\begin{aligned}\langle X, \nu_E^{eu}\rangle &= \langle X, \frac{\nabla f}{|\nabla f|}\rangle \\ &= \frac{\langle X, \nabla f\rangle}{|\nabla f|} = \frac{Xf}{|\nabla f|}.\end{aligned}$$

Thus

$$X\chi_E = -\frac{Xf}{|\nabla f|}\mathscr{H}^{n-1} \llcorner_{\partial E}. \tag{4.5.8}$$

4.6 Exercises

Exercise 4.6.1. *Let V and $W \subset \mathfrak{g}$ be two sub-vector spaces with $X_1, \dots, X_l$ and $Y_1, \dots, Y_m$ basis of V and W respectively. Then show that the vectors $[X_i, Y_j]$, for $i = 1, \dots, l, j = 1, \dots, m$ span $[V, W]$, thus one can extract a basis among such brackets.*

Exercise 4.6.2. *Let $\mathfrak{g} = V_1 \oplus \cdots \oplus V_s$ be a stratification of a Lie algebra. Assume that $X_{m_j+1}, \dots, X_{m_j}$ is a basis of V_j, then show that the order-reversed basis $X_n, \dots, X_1$ is a (strong) Malcev basis.*

Exercise 4.6.3. *Considering the horizontal path constructed in the proof of Property 3 in Proposition 4.1.6, give a lower bound on $d_{CC}(e, E(\boldsymbol{t}))$.*

Exercise 4.6.4. *Let G be a simply connected nilpotent Lie group and let V_1 be a sub-space such that*

$$\mathfrak{g} = V_1 \oplus [\mathfrak{g}, \mathfrak{g}].$$

Denote by $\mathfrak{g}^{(i)}$ the i-th term in the lower (or descending) central series of $\mathfrak{g}$, Show first that

$$\mathfrak{g}^{(2)} = [V_1, V_1] + \mathfrak{g}^{(3)}.$$

Then, by induction, show

$$\mathfrak{g}^{(i)} = [V_1, [V_1, [\dots, [V_1, V_1] \dots]] + \mathfrak{g}^{(i+1)},$$

where in the above bracket there are i many V_1's. Finally deduce that such a V_1 generates the whole Lie algebra.

Chapter 5

Nilpotent discrete groups

5.1 Elements of Geometric Group Theory

A *discrete* group Γ is a topological group that as topological space is discrete.

A set S inside a group Γ is said to be *generating* if there is no proper subgroup of Γ containing S. In other words, every element in the group Γ can be written as a finite product of elements in S. If one interprets the elements in S as words of an alphabet, then one can use the expression: 'each element in Γ is represented by a word with letters in S'.

A group is said to be *finitely generated* if it admits a finite generating set.

After having fixed such a set S, one can construct a geometric graph related to the group Γ.

Definition 5.1.1 (Cayley graph)**.** *Let Γ be a discrete group and let S be a generating set. The* (colored and directed) Cayley graph $\mathcal{G} = \mathcal{G}(\Gamma, S)$ *is the colored directed graph constructed as follows: The vertex set* Vertex($\mathcal{G}$) *of $\mathcal{G}$ is identified with Γ. Each generator s of S determines a color c_s and the directed edges of color c_s consists of the pairs of the form (g, gs), with $g \in \Gamma$.*

Geometric Group Theory mostly studies finitely generated groups considering the large scale geometry (or coarse geometry) of the Cayley graph. In such case, the set S is usually assumed to be finite, symmetric, i.e., $S = S^{-1}$, and not containing the identity element of the group. In this case, the (uncolored) Cayley graph is an ordinary graph: its edges are not oriented and it does not contain loops.

Definition 5.1.2 (Word metric)**.** *Let Γ be a discrete group and let S be a generating set. For any two elements g and $h \in \Gamma$, their* word distance *with respect to S, is denoted by $d_S(g, h)$ and is defined as the minimum number of elements (=letters) in S whose product (=word) equals $g^{-1}h$.*

Analogously, the word metric d_S *on the whole Cayley graph* $\mathcal{G}(\Gamma, S)$ *is the length metric that gives length* 1 *to each edge of* $\mathcal{G}(\Gamma, S)$*. We have then an isometry between* (Γ, d_S) *and the vertex set of the graph* $(\mathrm{Vertex}\,(\mathcal{G}(\Gamma, S))\,, d_S)$

The group Γ acts naturally on its Cayley graph $\mathcal{G}(\Gamma, S)$ sending the vertex h to the vertex gh, for each fixed $g \in \Gamma$. One can easily check that such left translations preserve the graph structure of $\mathcal{G}$.

Proposition 5.1.3 (Isometry of the left action)**.** *The left translation of a group* Γ *are isometries with respect to the word metric. Analogously, the left translations induce an isometric action of the group* Γ *on the metric space* $(\mathcal{G}(\Gamma, S), d_S)$*, and such action is transitive on the vertex set.*

The word metric on a group Γ is not unique, because different symmetric generating sets give different word metrics. However, finitely generated word metrics are unique up to biLipschitz equivalence.

Proposition 5.1.4 (Bilipschitz invariants of a group)**.** *If* S *and* S' *are two symmetric, finite generating sets for* Γ *with corresponding word metrics* d_S *and* $d_{S'}$*, then there is a constant* K *such that the identity map from* (Γ, d_S) *to* $(\Gamma, d_{S'})$ *is a* K*-biLipschitz map. In fact,* K *is just the maximum of the* d_S *word norms of elements of* S' *and the* $d_{S'}$ *word norms of elements of* S*.*

Definition 5.1.5 (Quasi-isometry)**.** *Suppose* (M_1, d_1) *and* (M_2, d_2) *are metric spaces, and* $f : M_1 \to M_2$ *is a function (not necessarily continuous). Then* f *is called a* (A, L)- quasi-isometric embedding, *with* $L \geq 1$ *and* $A \geq 0$*, if*

$$\frac{1}{L}\, d_2(f(x), f(y)) - A \leq d_1(x, y) \leq L\, d_2(f(x), f(y)) + A \quad \textit{for all} \quad x, y \in M_1.$$

Moreover, a quasi-isometric embedding is called a quasi-isometry *if there exists a constant* $C \geq 0$ *such that to every* $u \in M_2$ *there exists* $x \in M_1$ *with*

$$d_2(u, f(x)) \leq C.$$

The spaces M_2 *and* M_2 *are called* quasi-isometric *if there exists a quasi-isometry between them.*

Theorem 5.1.6 (Fundamental observation of Geometric Group Theory)**.** *Let* X *be a metric space which is geodesic and proper, let* Γ *be a group acting on* X *by isometries. Assume that the action is proper and the quotient space* X/Γ *is compact.*

Then the group Γ is finitely generated and quasi-isometric to X.

More precisely, for any $x_0 \in X$, the orbit mapping

$$\Gamma \to X$$

$$\gamma \mapsto \gamma(x_0)$$

is a quasi-isometry.

Such fact was known in the 50's. A proof can be essentially re-contruct from [Lemma 2]Milnor. A detailed proof is in [Theorem 23]delaharpe.

From the above fundamental observation we deduce that Geometric Group Theory links the study of fundamental groups of compact manifolds and their Riemnnian universal covers. Namely, let M be a compact differentiable manifold. Let $\pi_1(M)$ the fundamental group of M. By the above observation, such discrete group is finitely generated. We endow the group with a word metric. Fix now a Riemannian metric g on M. Then there is a unique Riemannian metric $\tilde{g}$ on the universal cover $\tilde{M}$ of M such that the universal projection

$$(\tilde{M}, \tilde{g}) \twoheadrightarrow (M, g)$$

is a local isometry. We refer to such $\tilde{g}$ as the lifted Riemannian metric. The crucial result is that the coarse geometry of $\tilde{M}$ is the same that the coarse geometry of $\pi_1(M)$. A prove of the following proposition can be found in the lecture notes of M. Kapovich on GGT, use his Lemma 1.31.

Proposition 5.1.7. *Assume M is a Riemannian manifold that is compact.*

(i) *The fundamental group $\pi_1(M)$ is finitely generated.*

(ii) *The universal cover $\tilde{M}$, endowed with the lifted Riemannian distance, is quasi-isometric to $\pi_1(M)$, endowed with any word metric.*

Proposition 5.1.8. *Assume G is a finitely generated group and $H < G$ a subgroup.*

(i) *If H has finite index in G, then G and H are quasi-isometric.*

(ii) *If H is a finite group (and it is normal in G), then G and G/H are quasi-isometric.*

Definition 5.1.9. *We say that a group G is* virtually nilpotent *if there exists a sub-group $H < G$ of finite index in G that is nilpotent.*

5.2 The growth rate of balls

The bilipschitz equivalence of word metrics implies in turn that the growth rate of a finitely generated group is a well-defined isomorphism invariant of the group Γ, independent of the choice of a finite generating set S. This implies in turn that various properties of growth, such as polynomial growth, the degree of polynomial growth, and exponential growth, are isomorphism invariants of groups.

Given a finitely generated group Γ, we fix a finite symmetric generating set S. For each $R > 0$, let $B_S(e, R)$ be the metric ball in Γ with respect the distance d_S with center the origin e and radius R. We then denote by $\#(B_S(e, R))$ the cardinality of the finite set $B_S(e, R)$.

Definition 5.2.1. *The* growth rate *of a finitely generated group* Γ *is the growth rate of the function* $R \mapsto \#(B_S(e, R))$.

5.2.1 Invariance of the growth rate

Proposition 5.2.2. *If two metric spaces are quasi isometric, then they have the same growth rate.*

Corollary 5.2.3. *Assume M is a Riemannian manifold that is compact. Then the grow rate of the group $\pi_1(M)$ is the same as the grow rate of the volume function on the universal cover of M.*

Namely, consider the Riemmanian structure on $\tilde{M}$ lifted from the structure on M. Let $\tilde{B}(p, r)$ be the metric ball in $\tilde{M}$. Let $\mathrm{vol}_{\tilde{M}}$ be the Riemmanian volume form on $\tilde{M}$. Then the above corollary states that there exist constants k, c such that, for all $R > 1$, one has the bounds

$$k^{-1}\ \#(B_S(e, c^{-1}R)) \leq \mathrm{vol}_{\tilde{M}}(\tilde{B}(p, R)) \leq k\ \#(B_S(e, cR)).$$

Now, if a group Γ is virtually nilpotent, then by definition it has a nilpotent sub-group Γ' of finite index. Then Γ and Γ' are quasi-isometric and thus have the same growth rate. We will describe the fact that the groups that are virtually nilpotent are exactly those that have a polynomial growth rate.

5.2.2 Polynomial growth and virtual nilpotency

Definition 5.2.4 (Polynomial growth)**.** *A discrete group* Γ *is said to have* polynomial growth *if, for some (and thus for any) generating set S, there exist $C > 0$ and $k > 0$ such that for any integer $R \geq 1$*

$$\#(B_S(e, R)) \leq C \cdot R^k.$$

Another choice for S would only change the constant C, but not the polynomial nature of the bound, because of Proposition 5.2.2. Actually one only requires that the growth of the balls are bounded by a polynomial function. However, a result of Pansu states that, in fact, the above equation can be improved saying that there exists $c(S) > 0$ and an integer $d(\Gamma) \geq 0$ depending on Γ only such that the following holds:

$$\#(B_S(e, R)) = c(\Gamma)R^{d(\Gamma)} + o(R^{d(\Gamma)}), \qquad \text{as } R \to \infty.$$

The condition of polynomial growth can be further weakened, cf. [vdDW84, Kle09].

A result of J. Wolf is that a group has polynomial growth if it is nilpotent. A deep result of Gromov is the equivalence of polynomial growth and virtual nilpotency.

Theorem 5.2.5 (Gromov's polynomial growth)**.** *A finitely generated group has polynomial growth rate if and only if it is virtually nilpotent.*

The original proof in [Gro81] is based on Gleason-Montgomery-Zippin-Zippin-Yamabe structure theory of locally compact groups. A new short proof has been given by Kleiner in [Kle09].

A non trivial consequence of Gromov's Theorem is that if a group has polynomial growth then the exponent of the growth rate is an integer. The plan of this chapter is to give an exposition of how sub-Riemannian geometry plays a role in the polynomial growth theorem and observe that such integer exponent is in fact the Hausdorff dimension of a Carnot group associated to the finitely generated group.

5.3 Asymptotic cone

Theorem 5.3.1 (Wolf-Bass-Gromov-Pansu)**.** *The degree of growth of a finitely generated group Γ of polynomial growth is an integer and equals the Hausdorff dimension of the Carnot group that is the asymptotic cone of Γ.*

The asymptotic cone is also known as the tangent cone at infinity.

Theorem 5.3.2 (Pansu [Pan83b])**.** *The asymptotic cone of a nilpotent finitely generated discrete group Γ is a Carnot group G_∞ endowed with a left-invariant sub-Finsler structure. The Hausdorff dimension of G_∞ is the exponent of the growth rate of Γ.*

5.4 The Malcev closure

Briefly, a lattice is a discrete subgroup with finite covolume. Here is the formal definition:

Definition 5.4.1 (Lattice)**.** *Let G be a locally compact topological group. A subgroup $\Gamma < G$ is a* lattice *if it is discrete (as topological subspace) and has the property that on the quotient space G/Γ there is a finite G-invariant*[1] *measure.*

Proposition 5.4.2. *Let G be a Lie group endowed with a left-invariant Riemannian metric. Let Γ be a lattice in G. Then the quotient G/Γ is in fact compact and thus Γ is quasi-isometric to G.*

Theorem 5.4.3 ([Rag72, Theorem 2.18])**.** *A group Γ is isomorphic to a lattice in a simply connected nilpotent Lie group if and only if*

1. *Γ is finitely generated,*

2. *Γ is nilpotent, and*

3. *Γ has no torsion.*

Corollary 5.4.4 (Malcev Theorem [Mal51])**.** *If Γ is a finitely generated group which is nilpotent and has no torsion then it is isomorphic to a discrete cocompact subgroup of a simply connected nilpotent Lie group G.*

Some useful facts:

1. Every subgroup of a nilpotent group is nilpotent. (easy!)

2. Every subgroup of a finitely generated nilpotent group is finitely generated, cf. [Theorem 9.16]Mac or [Rag72, Theorem 2.7].

3. Every nilpotent group generated by finitely many elements of finite order is finite, cf. [Theorem 9.17]Mac.

These facts implies the following:

Lemma 5.4.5. *The elements of finite order in a nilpotent group G form a normal sub-group $Tor(G)$, called the torsion sub-group of G. If G is finitely generated, $Tor(G)$ is finite. The quotient $G/\mathrm{Tor}(G)$ is torsion-free, that is, its only element of finite order is the identity.*

Proposition 5.4.6. *Let Γ be a finitely generated discrete group Γ of polynomial growth, then Γ is quasi-isometric to a connected, simply connected, and nilpotent Lie group G.*

[1] Recall that the quotient on G/Γ is on the right, so G acts naturally on the left.

If a group Γ has polynomial growth, then, by Gromov Theorem 5.2.5, there is a subgroup $\Gamma_1 < \Gamma$ that is nilpotent and $[\Gamma, \Gamma_1] < \infty$. Let $\mathrm{Tor}(\Gamma_1)$ be the torsion of Γ_1, which is a *finite* and *normal* subgroup. Define $\Gamma_2 := \Gamma_1/\mathrm{Tor}(\Gamma_1)$. Then Γ_2 is nilpotent and has no torsion, thus, by Malcev Theorem 5.4.4, there is a connected, simply connected, and nilpotent Lie group G and a discrete cocompact subgroup $\Gamma' < G$, such that Γ_2 is isomorphic to Γ'.

The groups Γ, Γ_1, Γ_2, Γ', and G are quasi-isometric.

5.5 The graded algebra

Definition 5.5.1 (Graded algebra)**.** *Let $\mathfrak{g}$ be a Lie algebra that is nilpotent of step s. Let $\mathfrak{g}^{(i+1)} := [\mathfrak{g}; \mathfrak{g}^{(i)}]$ be the descending central series of $\mathfrak{g}$. The graded algebra of $\mathfrak{g}$ is the Lie algebra $\mathfrak{g}_\infty$ given by the direct sum decomposition*

$$\mathfrak{g}_\infty := \bigoplus_{i=1}^{s} \mathfrak{g}^{(i)}/\mathfrak{g}^{(i+1)},$$

endowed with the unique Lie bracket $[\cdot,\cdot]_\infty$ that has the property that, if $x \in \mathfrak{g}^{(i)}$ and $y \in \mathfrak{g}^{(j)}$, the bracket is defined, modulo $\mathfrak{g}^{(i+j)}$, as

$$[\bar{x}, \bar{y}]_\infty = \overline{[x, y]}.$$

5.6 The limit CC metric

Let Γ' be a discrete cocompact sub-group in a connected, simply connected, and nilpotent Lie group G. Let G_∞ be the unique connected, simply connected Lie group whose Lie algebra is the graded algebra $\mathfrak{g}_\infty$ of $\mathfrak{g}$.

Let $\|\cdot\| := d_S(e, \cdot)$ be a 'norm' on Γ' induced by a finite generating set S. We shall describe the CC metric induced on the Carnot group G_∞.

Consider the two sets:

$$A := \Gamma'/[\Gamma', \Gamma'] \qquad \text{and} \qquad B := G/[G, G].$$

Both A and B are Abelian groups. Moreover, B is a (finite dimensional) vector space.

A is a subgroup of B. (?!?)

$\|\cdot\|$ induces a norm on A. (?!?)

One defines

$$\|a\|_\infty := \lim_{k\to\infty} \frac{1}{k} \|ka\|.$$

Such norm extends to B. (?!?)

Recall that, as in any Carnot group, $V_1 \simeq \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$. Thus we consider the projection

$$\pi : G_\infty \twoheadrightarrow G_\infty/[G_\infty, G_\infty].$$

Therefore we can transport the norm on V_1, using the isomorphism between V_1 and $B := G/[G, G]$. (?!?)

Appendix A

Curves in sub-Finsler nilpotent groups

A.0.1 A special sub-Finsler geometry on nilpotent groups

Let G be a simply connected nilpotent Lie group. Let $V_1 \subseteq T_eG$ be a sub-vector space. Let Δ be the left-invariant distribution with $\Delta_e = V_1$. Considering V_1 as a sub-space of the Lie algebra $\mathfrak{g}$ of G, assume that the algebra generated by V_1 is the whole of $\mathfrak{g}$. In other words, assume that Δ is bracket generating. Thus we have the flag of left-invariant bundles

$$\Delta = \Delta^{[1]} \subseteq \Delta^{[2]} = \Delta + [\Delta, \Delta] \subseteq \ldots \subseteq \Delta^{[s]} = TG.$$

There is a one-to-one correspondence between vectors in V_1 and vector fields in the intersection $\mathfrak{g} \cap \Gamma(\Delta)$ of the Lie algebra of G and the sections of Δ. We will confuse the two notions without problems.

Fix a norm $\|\cdot\|$ on V_1. It extends to a left-invariant norm on Δ. The triple $(G, \Delta, \|\cdot\|)$ is a sub-Finsler manifold.

In the sequel, whenever we speak of the FCC metric on the simply connected nilpotent Lie group G, we mean one that is associated to a norm $\|\cdot\|$ on a sub-space V_1 such that $\mathfrak{g} = V_1 \oplus [\mathfrak{g}, \mathfrak{g}]$ where $\mathfrak{g} = Lie(G)$.

One can easily check that any such V_1 generates the Lie algebra, cf. Exercise 4.6.4.

Assume that G is a simply connected nilpotent Lie group with a left-invariant distribution Δ such that

$$\mathfrak{g} = \Delta_e \oplus [\mathfrak{g}, \mathfrak{g}], \tag{A.0.1}$$

as, for example, a Carnot group.

Question A.0.2. *If G is a simply connected nilpotent Lie group and V_1 and W_1 are sub-spaces such that*

$$\mathfrak{g} = V_1 \oplus [\mathfrak{g}, \mathfrak{g}] = W_1 \oplus [\mathfrak{g}, \mathfrak{g}],$$

then, does exist a Lie algebra isomorphism $\phi : \mathfrak{g} \to \mathfrak{g}$ such that $\phi(V_1) = W_1$? The aswer should be no, however, see Exercise 4.6.4.

Definition A.0.3 (The projection π_1)**.** *Let* $\mathrm{proj} : T_eG \to V_1 = \Delta_e$ *be the projection onto V_1 with kernel $[\mathfrak{g}, \mathfrak{g}]$. Define*

$$\begin{aligned} \pi_1 : \quad G &\to \quad V_1 \\ p &\mapsto \quad \pi_1(p) := \mathrm{proj}(\exp^{-1}(p)). \end{aligned} \tag{A.0.4}$$

Lemma A.0.5. *The following properties hold:*

(i) *The map $\pi_1 : (G, \cdot) \to (V_1, +)$ is a group homomorphism.*

(ii) *The differential of π_1 is the identity when restricted to V_1:*

$$d\pi_1|_{V_1} = \mathrm{id}_{V_1}.$$

Proof of (i). By Theorem 3.1.13, since G is a simply connected and nilpotent, for all p and $q \in G$, exist X and $Y \in \mathfrak{g}$ such that $\exp(X) = p$ and $\exp(Y) = q$. Then, by BCH formula and assumption (A.0.1)

$$\begin{aligned} \pi_1(p \cdot q) &= \mathrm{proj}(\exp^{-1}(pq)) = \mathrm{proj}\left(\exp^{-1}(\exp(X)\exp(Y))\right) \\ &= \mathrm{proj}\left(X + Y + \frac{1}{2}[X, Y] + \dots\right) = \mathrm{proj}(X + Y). \end{aligned}$$

On the other hand,

$$\pi_1(p) + \pi_1(q) = \mathrm{proj}(\exp^{-1}(p)) + \mathrm{proj}(\exp^{-1}(q)) = \mathrm{proj}(\exp^{-1}(p) + \exp^{-1}(q)) = \mathrm{proj}(X + Y).$$

Proof of (ii). Since Theorem 3.1.6(iii), $d\pi_1|_{V_1} = d(\mathrm{proj}|_{V_1}) = d(\mathrm{id}\,|_{V_1}) = \mathrm{id}_{V_1}$. □

The "development" of a curve

The map π_1 is useful since it gives a second link between the tangents of an horizontal curves and vector at the identity. Let $\gamma(t)$ be an absolute continuous curve with $\dot{\gamma}(t)$ horizontal, i.e.,

$\dot{\gamma}(t) \in \Delta_{\gamma(t)} \subseteq T_{\gamma(t)}G$ for almost every t. The vector $\dot{\gamma}(t)$ can be identified with a vector in V_1, so as a tangent vector at the identity. We define $\gamma'(t) \in V_1 \subseteq T_eG$ as

$$\gamma'(t) := (L_{\gamma(t)})_*^{-1}\dot{\gamma}(t).$$

We then have the following formula

$$\gamma'(t) = \frac{d}{dt}(\pi_1 \circ \gamma)(t) \tag{A.0.6}$$

Proof of Formula (A.0.6). Using Lemma A.0.5, and that $\pi_1(e) = 0$, we get

$$\begin{aligned}
\frac{d}{dt}(\pi_1 \circ \gamma)(t) &= \lim_{h\to 0} \frac{\pi_1(\gamma(t+h)) - \pi_1(\gamma(t))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(\gamma(t)^{-1}) + \pi_1(\gamma(t+h))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(\gamma(t)^{-1}\gamma(t+h))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(L^{-1}_{\gamma(t)}\gamma(t+h))}{h} \\
&= \lim_{h\to 0} \frac{\pi_1(L^{-1}_{\gamma(t)}\gamma(t+h)) - \pi_1(L^{-1}_{\gamma(t)}\gamma(t))}{h} \\
&= \frac{d}{dh}\left((\pi_1 \circ L^{-1}_{\gamma(t)} \circ \gamma\right)(t+h)\bigg|_{h=0} \\
&= (\pi_1)_* \circ (L^{-1}_{\gamma(t)})_*\dot{\gamma}(t) \\
&= \mathrm{id}(\gamma'(t)) = \gamma'(t)
\end{aligned}$$

□

A.0.2 Horizontal lines as geodesics

Definition A.0.7. *Let $X \in V_1$. The curve $\gamma(t) := \exp(tX)$ is the one-parameter sub-group of the horizontal vector X, and it is called the horizontal line in the direction of X.*

The curve $\gamma(t)$ is obviously horizontal with respect to Δ, since

$$\dot{\gamma}(t) = X_{\gamma(t)} \in \Delta_{\gamma(t)}.$$

The length of $\gamma(t)$, for $t \in [0, T]$, with respect to the CC metric of $(M, \Delta, \|\cdot\|)$ is $T\|X\|$. Indeed,

$$\begin{aligned} \text{Length}(\gamma) &= \int_0^T \|\dot{\gamma}(t)\|\, dt \\ &= \int_0^T \left\|X_{\gamma(t)}\right\| dt \\ &= \int_0^T \left\|\left(L_{\gamma(t)}\right)_* X_e\right\| dt \\ &= \int_0^T \|X\|\, dt \\ &= T\|X\|, \end{aligned}$$

where we used that both X and the norm are left-invariant. Thus we get the formula

$$\text{Length}\left(\exp(tX)|_{t\in[0,T]}\right) = T\|X\|. \tag{A.0.8}$$

In a Lie group endowed with a left-invariant Riemannian metric, the one-parameter subgroups are NOT always geodesics.

For instance in $SL(2, \mathbb{R})$ the upper triangular unipotent one parameter subgroup

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

is not a geodesic, because it's distance to Id is roughly $\log(t)$, not t.

In a non-compact simple Lie group only the one-parameter groups coming from the p part of the Cartan decomposition will be geodesics.

Also in the Heisenberg group, if you consider the vertical line, then it is a one-parameter group, but not a geodesic in Riemannian left-invariant metrics.

Proposition A.0.9. *Let G be any Lie group endowed with a left-invariant Riemannian metric. Then the one-parameter subgroups in the direction of X is geodesic if and only if X is orthogonal to $[X, \mathfrak{g}]$.*

Proposition A.0.10. *Consider a nilpotent Lie group G endowed with a left-invariant sub-Finsler distance with respect to some distribution Δ such that*

$$\Delta \oplus [\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}.$$

Then one-parameter subgroups of horizontal vectors are geodesics.

Proof. What we need to show is that

$$\|\pi_1(g)\| \leq d_{CC}(e,g), \tag{A.0.11}$$

where d_{CC} is the Finsler-Carnot-Carathéodory distance and π_1 is the projection defined in (A.0.4).

Restricting (A.0.11) to g belonging to $\exp(V_1)$ will finish the proof because of calculation (A.0.8). Indeed, if now $X \in V_1$ then the curve $t \mapsto \exp(tX)$ is a geodesic since

$$d_{CC}(e,\exp(TX)) \leq \text{Length}\left(\exp(tX)|_{t\in[0,T]}\right) = T\,\|X\| = \|TX\| \leq d_{CC}(e,\exp(TX)).$$

Now inequality (A.0.11) is true because, by definition of the metric on G, there is a sequence of piece-wise linear (or piece-wise smooth) horizontal curves joining e and g whose length tends to $d_{CC}(e,g)$. But if $\gamma(t):[0,1]\to G$ is such a curve, then, by Formula (A.0.6),

$$\begin{aligned}
\|\pi_1(g)\| &= \|\pi_1(\gamma(1)) - \pi_1(\gamma(0))\| \\
&= \left\|\int_0^1 \frac{d}{dt}\left(\pi_1\circ\gamma\right)(t)dt\right\| \\
&\leq \int_0^1 \left\|\frac{d}{dt}\left(\pi_1\circ\gamma\right)(t)\right\| dt \\
&= \int_0^1 \|\gamma'(t)\|\,dt \\
&= \int_0^1 \left\|(L_{\gamma(t)})_*^{-1}\dot\gamma(t)\right\| dt \\
&= \int_0^1 \|\dot\gamma(t)\|\,dt \\
&= \text{Length}(\gamma).
\end{aligned}$$

□

A.0.3 Lifts of curves

Lemma A.0.12. *Let $\phi:\mathfrak{g}\to\mathfrak{h}$ a Lie algebra homomorphism.*

(i) *If ϕ has the property that $\phi(V_1^{(G)}) \subseteq V_1^{(H)}$, then*

$$\text{proj}^H \circ \phi = \phi \circ \text{proj}^G,$$

where $\text{proj}^G:\mathfrak{g}\to\mathfrak{g}$ and $\text{proj}^H:\mathfrak{h}\to\mathfrak{h}$ are the projections onto V_1^G and V_1^H respectively with kernels $[\mathfrak{g},\mathfrak{g}]$ and $[\mathfrak{h},\mathfrak{h}]$ respectively.

(ii) *If ϕ has the property that $\phi(V_1^{(G)}) \supseteq V_1^{(H)}$, then ϕ is surjective.*

(iii) *If*

$$\varphi_*|_{V_1^{(G)}} : V_1^{(G)} \to V_1^{(H)}$$

is an isometry of normed spaces, then $\varphi : G \to H$ is a 1-Lipschitz, with respect to the respective FCC metrics.

Proof of (i). If $X \in V_1^{(G)}$, then $(\phi \circ \mathrm{proj})(X) = \phi(X)$. Since by assumption we also have $\phi(X) \in V_1^{(H)}$, then $(\mathrm{proj} \circ \phi)(X) = \phi(X)$. So $\mathrm{proj} \circ \phi$ and $\phi \circ \mathrm{proj}$ are two homomorphisms that coincide on $V_1^{(G)}$. Since $V_1^{(G)}$ generates the algebra $\mathfrak{g}$, then the two homomorphisms are equal.

Proof of (ii). It is obvious since $\phi(\mathfrak{g})$ is an algebra that contains the generating sub-space $V_1^{(H)}$.

Proof of (iii). It is enough to observe that if $\gamma : [0,1] \to G$ is a geodesic, then

$$\begin{aligned} d(\gamma(0), \gamma(1)) &= \mathrm{Length}(\gamma) \\ &= \int_0^1 \|\dot{\gamma}(t)\| \, dt \\ &= \int_0^1 \|\gamma'(t)\|_{V_1(G)} \, dt \\ &= \int_0^1 \|\varphi_*(\gamma'(t))\|_{V_1(H)} \, dt \\ &= \int_0^1 \left\| \frac{d}{dt} (\varphi(\gamma(t))) \right\|_{V_1(H)} dt \\ &= \mathrm{Length}(\varphi \circ \gamma) \\ &\geq d(\varphi(\gamma(0)), \varphi(\gamma(1))). \end{aligned}$$

□

Lemma A.0.13. *The projection map $\pi_1 : G \to V_1$ has the following properties.*

(i) *For any Lipschitz curve σ in V_1 with $\sigma(0) = 0$, there exists a unique Lipschitz horizontal curve γ with $\pi_1(\gamma) = \sigma$ and $\gamma(0) = e$, and such a curve is the solution of the ODE*

$$\left\{ \begin{array}{lcl} \dot{\gamma}(t) &=& (L_{\gamma(t)})_* \dot{\sigma}(t) \\ \gamma(0) &=& e. \end{array} \right. \tag{A.0.14}$$

(ii) *The length of the horizontal curves equals the length of their projections:*

$$\mathrm{Length}(\gamma) = \mathrm{Length}(\pi_1 \circ \gamma),$$

for all horizontal curves γ, with $\gamma(0) = e$, where the first length is with respect to the FCC metric and the second one is in the normed space $(V_1, \|\cdot\|)$.

(iii) *If $\varphi : G \to H$ is a Lie group homomorphism with $\varphi_*(V_1^{(G)}) \subseteq V_1^{(H)}$, then*

$$\pi_1^{(H)} \circ \varphi \circ \gamma = \varphi_* \circ \pi_1^{(G)} \circ \gamma,$$

for all horizontal curves γ, with $\gamma(0) = e$.

Proof of (i). The existence of a solution of the ODE is a consequence of the general Carathéodory's theorem, cf. cite[page 43]Coddington-Levinson (1955). The uniquesess can be shown proving that, if $\gamma_1(t)$ and $\gamma_2(t)$ are two solutions, then

$$\frac{d}{dt}\left(\gamma_1(t)\gamma_2(t)^{-1}\right) \equiv 0.$$

Let $\gamma(t)$ be the solution of the ODE. Then

$$\gamma'(t) = (L_{\gamma(t)})^* \dot{\gamma}(t) = \frac{d}{dt}(\sigma(t)).$$

Since Formula (A.0.6), we have that $\pi_1 \circ \gamma$ and σ are two curves in V_1 with same starting point $\pi_1(\gamma(0)) = 0 = \sigma(0)$ and same derivative

$$\frac{d}{dt}(\pi_1 \circ \gamma) = \frac{d}{dt}\sigma.$$

Therefore $\pi_1 \circ \gamma = \sigma$.

Proof of (ii). By Formula (A.0.6), one has

$$\begin{aligned}
\text{Length}(\pi_\circ \gamma) &= \int_0^1 \left\| \frac{d}{dt}\left(\pi_1 \circ \gamma\right)(t) \right\| dt \\
&= \int_0^1 \|\gamma'(t)\|\, dt \\
&= \int_0^1 \left\|(L_{\gamma(t)})_*^{-1} \dot{\gamma}(t)\right\| dt \\
&= \int_0^1 \|\dot{\gamma}(t)\|\, dt \\
&= \text{Length}(\gamma).
\end{aligned}$$

Proof of (iii). By Theorem 3.1.7 and Lemma A.0.12(i), one has

$$\begin{aligned}
\pi_1^{(H)} \circ \varphi \circ \gamma &= \operatorname{proj} \circ \exp^{-1} \circ \varphi \circ \gamma \\
&= \operatorname{proj} \circ \varphi_* \circ \exp^{-1} \circ \gamma \\
&= \varphi_* \circ \operatorname{proj} \circ \exp^{-1} \circ \gamma \\
&= \varphi_* \circ \pi_1^{(G)} \circ \gamma.
\end{aligned}$$

□

We have the following formula combining (i) and (iii):

$$(d\varphi)_{\gamma(t)}\dot{\gamma}(t) = \left((L_{\varphi(\gamma(t))})_* \circ \varphi_*\right)\left(\frac{d}{dt}(\pi \circ \gamma)(t)\right).$$

A.0.4 Curves in free nilpotent Lie groups

Proposition A.0.15. *Assume that each pair of points in G can be joined by a smooth geodesic. If there is a homomorphism $\varphi : G \to H$ such that*

$$\varphi_*|_{V_1^{(G)}} : V_1^{(G)} \to V_1^{(H)}$$

is an isometry of normed spaces, then each pair of points in H can be joined by a smooth geodesic.

In the proposition, the word smooth can be replaced by C^k, C^ω, or piece-wise linear, since the good geodesics in H will be images under φ of good geodesics in G.

Proof. Now pick a point $p \in H$ and a geodesic $\xi : [0,1] \to H$ connecting the identity.to p. Then Push the curve on $V_1^{(H)}$ and then back to $V_1^{(G)}$, i.e., consider the curve

$$(\varphi_*)^{-1} \circ \pi_1 \circ \xi.$$

By Lemma A.0.13(i) consider a/the curve $\tilde{\xi}$ such that

$$\pi_1^{(G)} \circ \tilde{\xi} = (\varphi_*)^{-1} \circ \pi_1^{(H)} \circ \xi.$$

Note that $\varphi(\tilde{\xi}(1)) = p$. In fact $\varphi \circ \tilde{\xi} = \xi$. Indeed, $\varphi \circ \tilde{\xi}$ and ξ are the unique lift of $\pi_1^{(H)} \circ \xi$ under $\pi_1^{(H)}$, since

$$\pi_1^{(H)} \circ \varphi \circ \tilde{\xi} = \varphi_* \circ \pi_1^{(G)} \circ \tilde{\xi} = \varphi_* \circ (\varphi_*)^{-1} \circ \pi_1^{(H)} \circ \xi = \pi_1^{(H)} \circ \xi,$$

where we initially used Lemma A.0.13(iii). Let $\tilde{\gamma}$ be a smooth geodesic joining e to $\tilde{\xi}(1)$. We claim that $\varphi \circ \tilde{\gamma}$ is the desired geodesic. Indeed, using, in order, that φ is 1-Lipschitz (cf. Lemma A.0.12(iii)), Lemma A.0.13(ii), the assumption $\varphi_*|_{V_1^{(G)}}$ isometry, and Lemma A.0.13(ii) again, we get

$$\begin{aligned} L(\varphi \circ \tilde{\gamma}) &\leq L(\tilde{\gamma}) \\ &\leq L(\tilde{\xi}) \\ &= L(\pi_1^{(G)} \circ \tilde{\xi}) \\ &= L((\varphi_*)^{-1} \circ \pi_1^{(H)} \circ \xi) \\ &= L(\pi_1^{(H)} \circ \xi) \\ &= L(\xi) \\ &= d(e, p). \end{aligned}$$

□

Let G and H two nilpotent Lie groups, with horizontal layers $V_1^{(G)}$ and $V_1^{(H)}$, respectively. Consider a homomorphism

$$\varphi : G \to H,$$

such that

$$\varphi_*|_{V_1^{(G)}} : V_1^{(G)} \to V_1^{(H)}$$

and it is an isomorphism. Notice that such φ is surjective.

Endow just H with a (left-invariant) FCC-metric with $V_1^{(H)}$ as horizontal bundle. In other words, we have fixed a norm $\|\cdot\|_H$ on $V_1^{(H)}$.

Considering that $\varphi_*|_{V_1^{(G)}}$ is an isomorphism, we might consider the following norm $\|\cdot\|_G$ on $V_1^{(G)}$:

$$\|v\|_G := \|\varphi_*(v)\|_H \,, \qquad \text{for } v \in V_1^{(G)}.$$

Such a norm induces a (left-invariant) FCC-metric with $V_1^{(G)}$ as horizontal bundle.

Then our surjective homomorphism $\varphi : G \to H$ becomes 1-Lipschitz (cf. Lemma A.0.12(iii)).

Now suppose we know that the problem has a positive answer for G, i.e. that any point in G can be joined to the identity by a piece-wise linear geodesic.

Now pick a point in H and a geodesic connecting this point to the identity. This geodesic lifts to a rectifiable path in G with the same length (because of the choice of lifted FCC on G). Of course here I'm using the fact that since the homomorphism is surjective, the dimension of the abelianisation of M is at least that of N.

Now observe that this new path on G is a geodesic, because otherwise there would be another path joining the endpoints of strictly smaller length ; however its projection to H will also be of strictly smaller length, because we said out projection map was 1-Lipschitz, thus contradicting that we had started with a geodesic in H.

Ok, so now we have this lifted path in G and we know it's a geodesic. By assumption we may now find another geodesic, piece-wise linear this time, joining the two points. Then its projection will also be piece-wise linear of course and it will again be a geodesic because once again the projection is 1-Lipschitz.

A.0.5 Open questions

Question A.0.16. *If ρ is a FCC metric w.r.t. a polyhedral unit ball on G, then does there exist a constant K such that for any p and q there exists a geodesic for ρ joining p and q that has less than K breack points?*

Question A.0.17. *Let G be a free nilpotent Lie group. If ρ is a FCC metric w.r.t. a stricly convex unit ball on G, then, for any p and q, does there exist a smooth geodesic for ρ joining p and q?*

Question A.0.18. *Let G be a connected simply connected nilpotent Lie group. If ρ is a G-invariant metric which is coarsely geodesic, i.e.,*

$$d(x, y) \geq L(\gamma_{x,y}) + C.$$

Is ρ at bounded distance from a FCC metric.

Bibliography

[AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[AG01] Andrei Agrachev and Jean-Paul Gauthier. On the subanalyticity of Carnot-Caratheodory distances. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 18(3):359–382, 2001.

[AK00] Luigi Ambrosio and Bernd Kirchheim. Rectifiable sets in metric and Banach spaces. *Math. Ann.*, 318(3):527–555, 2000.

[AKLD08] Luigi Ambrosio, Bruce Kleiner, and Enrico Le Donne. Rectifiability of sets of finite perimeter in carnot groups: existence of a tangent hyperplane. *accepted for publication in the Journal of Geometric Analysis*, 2008.

[AP94] Marco Abate and Giorgio Patrizio. *Finsler metrics—a global approach*, volume 1591 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1994. With applications to geometric function theory.

[BBI01] Dmitriĭ Burago, Yuriĭ Burago, and Sergeiĭ Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BCS00] D. Bao, S.-S. Chern, and Z. Shen. *An introduction to Riemann-Finsler geometry*, volume 200 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[BDGG69] E. Bombieri, E. De Giorgi, and E. Giusti. Minimal cones and the Bernstein problem. *Invent. Math.*, 7:243–268, 1969.

[Bel96] André Bellaïche. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 1–78. Birkhäuser, Basel, 1996.

[Bre07] Emmanuel Breuillard. Geometry of locally compact groups of polynomial growth and shape of large balls. *Preprint*, 2007.

[Bul02] Marius Buliga. Sub-riemannian geometry and lie groups. part i. *Seminar Notes, DMA-EPFL, Preprint on Arxiv*, 2002.

[CCM08] Luca Capogna, Giovanna Citti, and Maria Manfredini. Regularity of minimal surfaces in the one-dimensional Heisenberg group. In *"Bruno Pini" Mathematical Analysis Seminar, University of Bologna Department of Mathematics: Academic Year 2006/2007 (Italian)*, pages 147–162. Tecnoprint, Bologna, 2008.

[CG90] Lawrence J. Corwin and Frederick P. Greenleaf. *Representations of nilpotent Lie groups and their applications. Part I*, volume 18 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990. Basic theory and examples.

[CHY07] Jih-Hsin Cheng, Jenn-Fang Hwang, and Paul Yang. Existence and uniqueness for p-area minimizers in the Heisenberg group. *Math. Ann.*, 337(2):253–293, 2007.

[CHY09] Jih-Hsin Cheng, Jenn-Fang Hwang, and Paul Yang. Regularity of C^1 smooth surfaces with prescribed p-mean curvature in the Heisenberg group. *Math. Ann.*, 344(1):1–35, 2009.

[CJT06] Y. Chitour, F. Jean, and E. Trélat. Genericity results for singular curves. *J. Differential Geom.*, 73(1):45–73, 2006.

[DG54] Ennio De Giorgi. Su una teoria generale della misura $(r-1)$-dimensionale in uno spazio ad r dimensioni. *Ann. Mat. Pura Appl. (4)*, 36:191–213, 1954.

[DG55] Ennio De Giorgi. Nuovi teoremi relativi alle misure $(r-1)$-dimensionali in uno spazio ad r dimensioni. *Ricerche Mat.*, 4:95–113, 1955.

[DGN08] D. Danielli, N. Garofalo, and D. M. Nhieu. A notable family of entire intrinsic minimal graphs in the Heisenberg group which are not perimeter minimizing. *Amer. J. Math.*, 130(2):317–339, 2008.

[Fed69] Herbert Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[FSSC96] Bruno Franchi, Raul Serapioni, and Francesco Serra Cassano. Meyers-Serrin type theorems and relaxation of variational integrals depending on vector fields. *Houston J. Math.*, 22(4):859–890, 1996.

[FSSC01] Bruno Franchi, Raul Serapioni, and Francesco Serra Cassano. Rectifiability and perimeter in the Heisenberg group. *Math. Ann.*, 321(3):479–531, 2001.

[FSSC03] Bruno Franchi, Raul Serapioni, and Francesco Serra Cassano. On the structure of finite perimeter sets in step 2 Carnot groups. *J. Geom. Anal.*, 13(3):421–466, 2003.

[GN96] Nicola Garofalo and Duy-Minh Nhieu. Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. *Comm. Pure Appl. Math.*, 49(10):1081–1144, 1996.

[Goo76] Roe W. Goodman. *Nilpotent Lie groups: structure and applications to analysis.* Lecture Notes in Mathematics, Vol. 562. Springer-Verlag, Berlin, 1976.

[Gro81] Mikhael Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.

[Gro99] Mikhail Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1999. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[Hei01] Juha Heinonen. *Lectures on analysis on metric spaces.* Universitext. Springer-Verlag, New York, 2001.

[Jac79] Nathan Jacobson. *Lie algebras.* Dover Publications Inc., New York, 1979. Republication of the 1962 original.

[Kle09] Bruce Kleiner. A new proof of gromov's theorem on groups of polynomial growth. *Preprint*, 2009.

[LDM08] Enrico Le Donne and Valentino Magnani. Measure of submanifolds in the engel group. *accepted for publication in the Revista Matematica Iberoamericana*, 2008.

[LM08] Gian Paolo Leonardi and Roberto Monti. End-point equations and regularity of sub-Riemannian geodesics. *Geom. Funct. Anal.*, 18(2):552–582, 2008.

[LR03] G. P. Leonardi and S. Rigot. Isoperimetric sets on Carnot groups. *Houston J. Math.*, 29(3):609–637 (electronic), 2003.

[Mag08a] Valentino Magnani. Blow-up estimates at horizontal points and applications. *Preprint*, 2008.

[Mag08b] Valentino Magnani. Non-horizontal submanifolds and coarea formula. *J. Anal. Math.*, 106:95–127, 2008.

[Mal51] A. I. Malcev. On a class of homogeneous spaces. *Amer. Math. Soc. Translation*, 1951(39):33, 1951.

[MM95] Gregori A. Margulis and George D. Mostow. The differential of a quasi-conformal mapping of a Carnot-Carathéodory space. *Geom. Funct. Anal.*, 5(2):402–433, 1995.

[Mon02] Richard Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[MR09] Roberto Monti and Matthieu Rickly. Convex isoperimetric sets in the Heisenberg group. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 8(2):391–415, 2009.

[MV08] Valentino Magnani and Davide Vittone. An intrinsic measure for submanifolds in stratified groups. *J. Reine Angew. Math.*, 619:203–232, 2008.

[Pan82] Pierre Pansu. Une inégalité isopérimétrique sur le groupe de Heisenberg. *C. R. Acad. Sci. Paris Sér. I Math.*, 295(2):127–130, 1982.

[Pan83a] P. Pansu. An isoperimetric inequality on the Heisenberg group. *Rend. Sem. Mat. Univ. Politec. Torino*, (Special Issue):159–174 (1984), 1983. Conference on differential geometry on homogeneous spaces (Turin, 1983).

[Pan83b] Pierre Pansu. Croissance des boules et des géodésiques fermées dans les nilvariétés. *Ergodic Theory Dynam. Systems*, 3(3):415–445, 1983.

[Pan89] Pierre Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[Pau04] Scott D. Pauls. Minimal surfaces in the Heisenberg group. *Geom. Dedicata*, 104:201–231, 2004.

[Pau06] Scott D. Pauls. H-minimal graphs of low regularity in $\mathbb{H}^1$. *Comment. Math. Helv.*, 81(2):337–381, 2006.

[Rag72] M. S. Raghunathan. *Discrete subgroups of Lie groups*. Springer-Verlag, New York, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68.

[Rit09] Manuel Ritoré. Examples of area-minimizing surfaces in the sub-Riemannian Heisenberg group $\mathbb{H}^1$ with low regularity. *Calc. Var. Partial Differential Equations*, 34(2):179–192, 2009.

[RR08] Manuel Ritoré and César Rosales. Area-stationary surfaces in the Heisenberg group $\mathbb{H}^1$. *Adv. Math.*, 219(2):633–671, 2008.

[SCP08] Alessandro Sarti, Giovanna Citti, and Jean Petitot. The symplectic structure of the primary visual cortex. *Biol. Cybernet.*, 98(1):33–48, 2008.

[Sem96] Stephen Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.

[Ste74] P. Stefan. Accessible sets, orbits, and foliations with singularities. *Proc. London Math. Soc. (3)*, 29:699–713, 1974.

[Str86] Robert S. Strichartz. Sub-Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.

[Str89] Robert S. Strichartz. Corrections to: "Sub-Riemannian geometry" [J. Differential Geom. **24** (1986), no. 2, 221–263; MR0862049 (88b:53055)]. *J. Differential Geom.*, 30(2):595–596, 1989.

[Sus73] Héctor J. Sussmann. Orbits of families of vector fields and integrability of distributions. *Trans. Amer. Math. Soc.*, 180:171–188, 1973.

[vdDW84] L. van den Dries and A. J. Wilkie. Gromov's theorem on groups of polynomial growth and elementary logic. *J. Algebra*, 89(2):349–374, 1984.

[War83] Frank W. Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1983. Corrected reprint of the 1971 edition.